

## INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

<b>(51) International Patent Classification <sup>5</sup> :</b>  <b>C12N 15/82, 5/10, A01H 5/00</b>	<b>A1</b>	<b>(11) International Publication Number:</b> <b>WO 91/09957</b>  <b>(43) International Publication Date:</b> 11 July 1991 (11.07.91)
<b>(21) International Application Number:</b> PCT/US90/07295 <b>(22) International Filing Date:</b> 19 December 1990 (19.12.90)  <b>(30) Priority data:</b> 455,221                      22 December 1989 (22.12.89)    US  <b>(60) Parent Application or Grant</b> <b>(63) Related by Continuation</b> US                                      455,221 (CIP) Filed on                              22 December 1989 (22.12.89)  <b>(71) Applicant (for all designated States except US):</b> E.I. DU PONT DE NEMOURS AND COMPANY [US/US]; 1007 Market Street, Wilmington, DE 19898 (US).		<b>(72) Inventors; and</b> <b>(75) Inventors/Applicants (for US only) :</b> ODELL, Joan, Tellef- son [US/US]; 127 Monitor Place, Unionville, PA 19375 (US). RUSSELL, Sandra, Hoff [US/US]; 9 Cook Court, Avondale, PA 19311 (US). SAUER, Brian, Lee [US/ US]; 2514 Redstart Court, Wilmington, DE 19805 (US). HSU, Francis, Chuoh [CN/US]; 434 Coldspring Run, Newark, DE 19711 (US).  <b>(74) Agents:</b> HAMBY, William, H. et al.; E.I. du Pont de Ne- mours and Company, Legal/Patent Records Center, 1007 Market Street, Wilmington, DE 19898 (US).  <b>(81) Designated States:</b> AT (European patent), AU, BE (Euro- pean patent), BR, CA, CH (European patent), DE (Eu- ropean patent), DK (European patent), ES (European patent), FR (European patent), GB (European patent), GR (European patent), IT (European patent), JP, LU (European patent), NL (European patent), SE (Euro- pean patent), SU, US.  <b>Published</b> <i>With international search report.</i>
<b>(54) Title:</b> SITE-SPECIFIC RECOMBINATION OF DNA IN PLANT CELLS  <b>(57) Abstract</b>  A method for producing site-specific recombination of DNA in plant cells. A first DNA sequence comprising a first <i>lox</i> site and a second DNA sequence comprising a second <i>lox</i> site are introduced into the cells. The <i>lox</i> sites are contacted with <i>Cre</i> to produce recombination. Also disclosed are related plasmids, transformed plant cells, and plants containing the transformed cells.		

BEST AVAILABLE COPY

**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	ES	Spain	MG	Madagascar
AU	Australia	FI	Finland	ML	Mali
BB	Barbados	FR	France	MN	Mongolia
BE	Belgium	GA	Gabon	MR	Mauritania
BF	Burkina Faso	GB	United Kingdom	MW	Malawi
BG	Bulgaria	GN	Guinea	NL	Netherlands
BJ	Benin	GR	Greece	NO	Norway
BR	Brazil	HU	Hungary	PL	Poland
CA	Canada	IT	Italy	RO	Romania
CF	Central African Republic	JP	Japan	SD	Sudan
CG	Congo	KP	Democratic People's Republic of Korea	SE	Sweden
CH	Switzerland	KR	Republic of Korea	SN	Senegal
CI	Côte d'Ivoire	LI	Liechtenstein	SU	Soviet Union
CM	Cameroon	LK	Sri Lanka	TD	Chad
CS	Czechoslovakia	LU	Luxembourg	TG	Togo
DE	Germany	MC	Monaco	US	United States of America
DK	Denmark				

TITLESITE-SPECIFIC RECOMBINATION OF DNA IN PLANT CELLS  
CROSS-REFERENCE TO RELATED APPLICATION

5        This application is a continuation-in-part of  
copenpending application Serial No. 07/455,221, filed  
December 22, 1989.

FIELD OF THE INVENTION

10       This invention relates to a method for producing  
site-specific recombination of DNA in plant cells and to  
the novel recombinant DNA constructs used to introduce  
and express the lox and cre components of the  
recombination system, as well as to transgenic lox and  
cre containing plants and their seeds.

15       BACKGROUND OF THE INVENTION

A variety of materials, systems and organisms have  
been the subject of genetic engineering to introduce  
systems to manipulate DNA.

20       Abremski et al., Cell, 32: 1301-1311 (1983) disclose  
a site-specific recombination system of bacteriophage  
P1. The system consists of a recombination site  
designated loxP and a recombinase designated Cre.  
Recombination between loxP sites on supercoiled, nicked-  
circle or linear DNA occurs in the presence of Cre.

25       Sauer, Molecular and Cellular Biology, 7: 2087-2096  
(1987) discloses that the loxP-cre recombination system  
functions in the yeast Saccharomyces cerevisiae. This  
system was used to excise a gene located between two lox  
sites which had been introduced into the yeast genome.

30       Cre was expressed from an inducible yeast GAL1 promoter  
and this cre gene was located on an autonomously  
replicating yeast vector.

35       Sauer and Henderson, Proc. Natl. Acad. Sci. USA, 85:  
5166-5170 (1988) disclose that the loxP-cre  
recombination system functions in a transient manner in

mouse cells in tissue culture. Cre was expressed from an inducible mouse metallothionein promoter, the cre gene being located on a papilloma virus replicon-containing vector. Excision of a gene located between two lox sites on a plasmid that was transiently introduced into cells, or of an insert in a gene of a herpesvirus vector, was demonstrated.

Sauer and Henderson, Nucl. Acids Res., 17: 147-161 (1989) disclose that the loxP-cre recombination system functions in a stably transformed mouse tissue culture cell line. Cre, expressed from a rous sarcoma virus promoter, caused excision of a gene located between two lox sites that were integrated in the mouse cell genome.

Gatz and Quail, Proc. Natl. Acad. Sci. USA, 85: 1394-1397 (1988) disclose that expression of the bacterial tet repressor protein in plant protoplasts in culture in a transient manner results in regulation of a CaMV 35S promoter that has tet operator sequences added to it and is also transiently present in the protoplasts.

Baker et al., Proc. Natl. Acad. Sci. USA, 83: 4844-4848 (1986) disclose that the controlling element called "activator" that is derived from maize can excise itself after being introduced into the tobacco genome. Lassner et al., Mol. Gen. Genet., 218: 25-32 (1989) disclose that the "activator" element can be separated into two functional components: i) an element with a large internal deletion that cannot excise itself, but can be excised by ii) an element with a terminal deletion that cannot excise. These two components were separately transformed into tomato plants, brought together by genetic crosses, and shown to result in excision of the first component in some cells. This experiment indicates that elements from one plant genome can lead to recombination in heterologous plant cells,



however the DNA sequences required for activity of the recombination site are not defined.

It is an object of the present invention to manipulate exogenous DNA once it is resident in the plant cell to enhance the ability to control trait expression in engineered plants. A feature of the present invention is the versatility of the method disclosed herein for producing site-specific recombination of DNA in plant cells in that the method is useful toward a wide variety of applications. These and other objects, features and advantages will become apparent upon having reference to the description of the invention herein.

#### SUMMARY OF THE INVENTION

The present invention provides a method for producing site-specific recombination of DNA in plant cells. The method (1) comprises:

- i) introducing into the cells a first DNA sequence comprising a first lox site, and a second DNA sequence comprising a second lox site, and
- ii) contacting the lox sites with Cre, thereby producing the site-specific recombination.

In a preferred embodiment, a third DNA sequence comprising a cre gene is also introduced into the cells. This third DNA sequence may further comprise a promoter that is active in plant cells and expression of the cre gene is produced by direction of the promoter. Another method of the present invention is directed to method (1), wherein the first and second DNA sequences are introduced into two different DNA molecules and the site-specific recombination is a reciprocal exchange of DNA segments proximate to the lox sites.

The present invention also provides a method of excising exogenous genes or DNA segments in transgenic plants. This method comprises:

5 1) introducing into the cells a DNA sequence comprising a first lox site, a second lox site in the same orientation as the first lox site, and a gene or a DNA sequence therebetween; and

10 2) contacting the lox sites with Cre, thereby excising the heterologous gene or DNA sequence. The gene may be an undesired marker or trait gene.

Further claimed herein are plant cells transformed with a DNA sequence comprising at least one lox site, or with a cre coding region. Various plants are also claimed herein, such as a plant containing cells  
15 transformed with a cre coding region, preferably having argonomic or horticultural utility. Plasmids are claimed, having at least one lox site, a pre-selected DNA segment selected from the group consisting of a gene, a coding region and a DNA sequence that influences  
20 gene expression in plant cells. Similarly, DNA sequences are claimed, such as the sequence comprising at least one lox site and a pre-selected DNA segment selected from the group consisting of a gene, a coding region, and a DNA sequence that influences gene  
25 expression in plant cells.

Typical trait genes of interest in the present invention include those encoding enzymes or other proteins to confer altered oil composition in seed; altered seed protein composition; altered carbohydrate  
30 composition in seed; altered carbohydrate composition in fruit; altered pollen development properties; herbicide resistance; fungicide resistance; insecticide resistance; and the like.

Typical marker genes include those conferring  
35 hygromycin resistance, kanamycin resistance, bleomycin

resistance, sulfonyleurea resistance, streptomycin resistance or phosphinothricin resistance; or  $\beta$ -glucuronidase.

The gene may cause disruption of the cells expressing it such as ones encoding an RNase, restriction endonuclease, protease, a ribozyme, or an antisense RNA.

DNA segments of interest include those that reduce or block expression of an adjacent gene such as a polyadenylation nucleotide sequence or one with ATG sequence(s) in it.

The DNA segment may influence plant gene expression, including but not limited to a polyadenylation nucleotide sequence; a promoter; a regulatory nucleotide sequence; a coding region; a ribozyme; and an antisense RNA sequence.

#### BRIEF DESCRIPTION OF FIGURES

The invention will be more fully appreciated and understood upon having reference to the following Figures.

Figure 1 shows maps of plasmids used in matings with Agrobacterium for plant transformations. Restriction sites used in making the constructions are marked as B: BamHI, C: ClaI, E: EcoRI, H: HindIII, P: PstI, S: Sall, X: XbaI.

(A) of this figure represents the Cre/Hpt-A plasmid.

(B) of this figure represents the Cre/Hpt-B plasmid.

(C) of this figure represents the loxP/NptII/Hra plasmid.

Figure 2 illustrates site-specific recombination in loxP plants re-transformed with Agrobacterium tumefaciens harboring the Cre/Hpt-B vector.

(A) is results from a callus induction assay.

(B) is a map of the lox region of the loxP/NptII/Hra vector.

(C) shows Southern blot analysis of re-transformed plants.

Figure 3 shows kanamycin resistance in loxP x Cre hybrids from homozygous parents.

5 Figure 4 shows a map of plasmid pZ4LoxAG that was introduced into Agrobacterium tumefaciens and then into plants.

10 Figure 5 shows a map of plasmid pBSCre103 that was introduced into Agrobacterium tumefaciens and then into plants.

#### DETAILED DESCRIPTION OF THE INVENTION

In the method of the invention using three DNA sequences, the first and second DNA sequences may be introduced into the cells connected by a pre-selected DNA segment. In such a case, the first and second lox sites may have the same orientation and the site-specific recombination of DNA is a deletion of the pre-selected DNA segment. The cre coding region may be derived from bacteriophage P1, and the first and second lox sites may be loxP or derivatives thereof. The pre-selected DNA segment is selected from the group consisting of a gene, a coding region, and a DNA sequence that influences gene expression in plant cells. Alternatively the segment may be an undesired marker or trait gene. The first and second lox sites may be selected to have opposite orientations and the site-specific recombination may be an inversion of the nucleotide sequence of the pre-selected DNA segment. In such case, the same selection of cre coding region, lox sites, and pre-selected DNA segment as referenced earlier are preferred. Similarly, in the aforementioned procedure wherein DNA sequences are introduced into two different DNA molecules, selections of cre coding region and lox sites as referenced earlier are also preferred.

35 Plant cells of the invention may contain Cre protein or

be transformed with a DNA sequence comprising at least one lox site. In the latter case, in such plants the DNA sequence may comprise two lox sites and a cre coding region, and preferably exhibit agronomic or horticultural utility. Plasmids, according to the invention, may have at least one lox site and a DNA sequence that influences gene expression in plant cells. For this type of plasmid, the DNA sequence may be a polyadenylation nucleotide sequence derived from the ribulose biphosphate carboxylase (Rubisco) small subunit gene. Alternatively, the DNA sequence is a promoter, or a regulatory nucleotide sequence. In the plasmid the DNA sequence may be a selection marker. Of particular interest is a plasmid having a cre coding region and a promoter that is active in plant cells. Particular plasmids of interest include Plasmid Cre/Hpt-A (characterized by the restriction enzyme map shown in Figure 1A, or a derivative thereof), Cre/Hpt-B (characterized by the restriction enzyme map shown in Figure 1B, or a derivative thereof), loxP/NptII/Hra (characterized by the restriction enzyme map shown in Figure 1C, or a derivative thereof) and pZ24loxAG (characterized by the restriction map shown in Figure 4, or a derivative thereof). Of particular interest is the use of the methods of the invention in the manufacture of seedless produce.

As used herein, the expression "site-specific recombination" is intended to include the following three events:

1. deletion of a pre-selected DNA segment flanked by lox sites,
2. inversion of the nucleotide sequence of a pre-selected DNA segment flanked by lox sites, and
3. reciprocal exchange of DNA segments proximate to lox sites located on different DNA molecules.

It is to be understood that this reciprocal exchange of DNA segments can result in an integration event.

In the context of this disclosure, a number of terms shall be utilized.

5       The expression "nucleotide sequence" refers to a polymer of DNA or RNA, which can be single- or double-stranded, optionally containing synthetic, non-natural, or altered nucleotides capable of incorporation into DNA or RNA polymers.

10       "DNA segment" refers to a linear fragment of single- or double-stranded deoxyribonucleic acid (DNA), which can be derived from any source. The expression "DNA in plant cells" includes all DNA present in plant cells. As used herein, a "gene" is intended to mean a DNA  
15       segment which is normally regarded as a gene by those skilled in the art.

"Coding region" refers to a DNA segment which encodes a regulatory molecule or any polypeptide.

20       The expression "regulatory molecule" refers to a polymer of ribonucleic acid (RNA), such as antisense RNA or a ribozyme, or a polypeptide which is capable of enhancing or inhibiting expression of a gene product.

"Gene product" refers to a polypeptide resulting from transcription, translation, and, optionally, post-  
25       translational processing of a selected DNA segment.

The term "expression" as used herein is intended to mean the translation to gene product from a gene coding for the sequence of the gene product. In expression, a DNA chain coding for the sequence of gene product is  
30       first transcribed to a complementary RNA which is called a messenger RNA and then, the thus transcribed messenger RNA is translated into the above-mentioned gene product.

As used herein, the term "promoter region" refers to a sequence of DNA, usually upstream (5') of the coding  
35       sequence, which controls the expression of the coding

region by providing the recognition for RNA polymerase and/or other factors required for transcription to start at the correct site. A "promoter fragment" constitutes a DNA sequence consisting of the promoter region.

5 A promoter region can include one or more regions which control the effectiveness of transcription initiation in response to physiological conditions, and a transcription initiation sequence.

"Tissue specific promoters" as referred to herein  
10 are those that direct gene expression primarily in specific tissues such as roots, leaves, stems, pistils, anthers, flower petals or epidermal layers. Transcription stimulators, enhancers or activators may be integrated into tissue specific promoters to create a  
15 promoter with a high level of activity that retains tissue specificity.

"Regulatory nucleotide sequence", as used herein, refers to a nucleotide sequence located proximate to a coding region whose transcription is controlled by the  
20 regulatory nucleotide sequence in conjunction with the gene expression apparatus of the cell. Generally, the regulatory nucleotide sequence is located 5' to the coding region. A promoter can include one or more regulatory nucleotide sequences.

25 "Polyadenylation nucleotide sequence" or "polyadenylation nucleotide region" refers to a nucleotide sequence usually located 3' to a coding region which controls the addition of polyadenylic acid to the RNA transcribed from the coding region in  
30 conjunction with the gene expression apparatus of the cell.

"DNA segment that influences gene expression in plant cells" can include a coding region, a promoter, a regulatory nucleotide sequence, a polyadenylation

nucleotide sequence, or other DNA sequence regarded as influencing gene expression by those skilled in the art.

As used herein, "transformation" means processes by which cells/tissues/plants acquire properties encoded on a nucleic acid molecule that has been transferred to the cell/tissue/plant. "Transferring" refers to methods to transfer DNA into cells including, but not limited to, microinjection, permeabilizing the cell membrane with various physical (e.g., electroporation) or chemical (e.g., polyethylene glycol, PEG) treatments, high-velocity microprojectile bombardment also termed biolistics, or infection with Agrobacterium tumefaciens or A. rhizogenes. As used herein, "transformant" means a plant which has acquired properties encoded on a nucleic acid molecule that has been transferred to cells during the process known as transformation. As used herein, "re-transformation" means transformation of cells/tissues/plants which are in themselves transformants.

As used herein, "sexual hybridization" means the production of offspring by crossbreeding of two plants that are genetically different, such as those which have different DNA sequences integrated into their genome.

As used within "integrated" means that the transferred DNA is incorporated into the plant genome.

As used herein the expression "lox site" means a nucleotide sequence at which the gene product of the cre gene, referred to herein as Cre, can catalyze a site-specific recombination. The loxP site is a 34 base pair nucleotide sequence which can be isolated from bacteriophage P1 by methods known in the art. One method for isolating a loxP site from bacteriophage P1 is disclosed by Hoess et al., Proc. Natl. Acad. Sci. USA, 79: 3398 (1982). The loxP site consists of two 13 base pair inverted repeats separated by an 8 base pair



spacer region. The nucleotide sequences of the inverted repeats and the spacer region are as follows:

ATAACTTCGTATA ATGTATGC TATACGAAGTTAT.

E. coli transformed with plasmid loxP/NptII/Hra carrying  
 5 two lox sites, one on either side of a polyadenylation  
 nucleotide sequence derived from a tobacco Rubisco small  
 subunit gene, has been deposited with the ATCC under the  
 Budapest treaty agreement and bears deposit accession  
 number 68177. This and other deposits are available to  
 10 the public upon the grant of a patent to the assignee.  
 However, it should be understood that the availability  
 of a deposit does not constitute a license to practice  
 the subject invention in derogation of patent rights  
 granted by governmental action. The lox sites and  
 15 intervening region can be excised from plasmid  
loxP/NptII/Hra with the restriction enzyme HindIII. In  
 addition, a preselected DNA segment can be inserted into  
loxP/NptII/Hra at the BamHI restriction enzyme site by  
 techniques known in the art. Other suitable lox sites  
 20 include loxB, loxL and loxR sites which are nucleotide  
 sequences isolated from E. coli. These sequences are  
 disclosed and described by Hoess et al., Proc. Natl.  
Acad. Sci. USA, 79: 3398 (1982). Lox sites can also be  
 produced by a variety of synthetic techniques which are  
 25 known in the art. For example, synthetic techniques for  
 producing lox sites are disclosed by Ito et al., Nuc.  
Acid Res., 10: 1755 (1982) and Ogilvie et al., Science,  
 214: 270 (1981).

Methods for introducing a DNA sequence into plant  
 30 cells are known in the art. Nucleic acids can generally  
 be introduced into plant protoplasts, with or without  
 the aid of electroporation, polyethylene glycol, or  
 other processes known to alter membrane permeability.  
 Nucleic acid constructs can also be introduced into  
 35 plants using vectors comprising part of the Ti- or Ri-

plasmid, a plant virus, or an autonomously replicating  
 sequence. Nucleic acid constructs can also be  
 introduced into plants by microinjection or by high-  
 velocity microprojectiles, also termed "particle  
 5 bombardment" or "biolistics" [Sanford, J. C., Tibtech  
 6: 299 (1988)], directly into various plant parts. The  
 preferred means of introducing a nucleic acid fragment  
 into plant cells involves the use of A. tumefaciens  
 containing the nucleic acid fragment between T-DNA  
 10 borders either on a disarmed Ti-plasmid (that is, a Ti-  
 plasmid from which the genes for tumorigenicity have  
 been deleted) or in a binary vector in trans to a  
 disarmed Ti-plasmid. The Agrobacterium can be used to  
 transform plants by inoculation of tissue explants, such  
 15 as stems, roots, or leaf discs, by co-cultivation with  
 plant protoplasts, or by inoculation of seeds or wounded  
 plant parts.

The range of crop species in which foreign genes can  
 be introduced is increasing rapidly as tissue culture  
 20 and transformation methods improve and as selectable  
 markers become available. Thus, this invention is  
 applicable to a broad range of agronomically or  
 horticulturally useful plants. The particular method  
 which is employed to introduce the DNA sequence into a  
 25 selected plant cell is not critical. In a preferred  
 embodiment, DNA sequences are introduced into plant  
 cells by co-cultivation of leaf discs with  
A. tumefaciens essentially as described by Horsch  
 et al., Science, 227: 1229-1231 (1985) omitting the  
 30 nurse cultures.

In the present method, the lox sites are contacted  
 with Cre, thereby producing the site specific  
 recombination. In one embodiment, Cre or cre messenger  
 RNA is introduced into the cells directly by  
 35 microinjection, biolistics, or other protein or RNA

introduction procedure. In a preferred embodiment, the cre coding region is introduced into the plant cell under the control of a promoter that is active in plant cells. Suitable regulatory nucleotide sequences are known in the art. The promoter which is employed with a selected plant cell is not critical to the method of the invention. A partial list of suitable promoters include the 35S promoter of cauliflower mosaic virus described by Odell et al., Nature, 313: 810-812 (1985); the promoter from the nopaline synthase gene of A. tumefaciens described by Depicker et al., J. of Mol. Appl. Genet., 1: 561-573 (1982); the promoter from a Rubisco small subunit gene described by Mazur and Chui, Nucleic Acids Research 13: 2373-2386 (1985); the 1' or 2' promoter from the TR-DNA of A. tumefaciens described by Velten et al., EMBO J. 12:2723-2730 (1984); the promoter of a chlorophyll a/b binding protein gene described by Dunsmuir et al., J. Mol. Appl. Genet. 2:285-300 (1983); the promoter of a soybean seed storage protein gene described by Chen et al., Proc. Natl. Acad. Sci. USA, 83: 8560-8564 (1986); and the promoter from the wheat EM gene described by Marcotte et al., Nature 335: 454-457 (1988). Cre can be expressed throughout the plant generally in all cells at all stages of development, or expression of cre can be more specifically controlled through the use of promoters or regulatory nucleotide sequences having limited expression characteristics. Cre can be expressed in a tissue specific manner, for example only in roots, leaves, or certain flower parts. Cre can be expressed in a developmentally specific time period, for example only during seed formation or during reproductive cell formation. Cre expression can also be placed under the control of a promoter that can be regulated by application of an inducer. In this case cre expression

is off or very low until the external inducer is applied. Promoters active in plant cells have been described that are inducible by heat shock [Gurley et al., Mol. Cell. Biol. 6: 559-565 (1986)], ethylene [Broglie et al., Plant Cell 1: 599-607 (1989)], auxin [Hagan and Guilfoyle, Mol. Cell. Biol. 5: 1197-1203 (1985)], abscisic acid [Marcotte et al., Nature 335: 454-457 (1988)], salicylic acid (EPO.332104A2 and EPO 337532A1), and substituted benzenesulfonamide safeners (WO 90/11361). Control of cre expression by the safener-inducible promoter 2-2, or its derivatives, allows the expression to be turned on only when the inducing chemical is applied and not in response to environmental or phytohormonal stimuli. Thus cre expression can be initiated at any desired time in the plant life cycle. Preferably, the regulatory nucleotide sequence is a 35S promoter or a 2-2 promoter.

The gene product of the cre coding region is a recombinase herein designated "Cre" which effects site-specific recombination of DNA at lox sites. As used herein, the expression "cre coding region" means a nucleotide sequence which codes for a gene product which effects site-specific recombination of DNA in plant cells at lox sites. One cre coding region can be isolated from bacteriophage P1 by methods known in the art. One method for isolating a cre coding region from bacteriophage P1 is disclosed by Abremski et al., Cell, 32: 1301-1311 (1983). The naturally occurring cre coding region can be altered by mutation to produce Cre proteins with altered properties as described by Wierzbicki et al., J. Mol. Biol., 195: 785-794 (1987). These altered Cre proteins retain their identities as Cre.

E. coli transformed with plasmid Cre/Hpt-A and E. coli transformed with plasmid Cre/Hpt-B both carrying a

cre coding region isolated from bacteriophage P1 and a cauliflower mosaic virus (CaMV) 35S promoter have been deposited with the ATCC and bear deposit accession numbers ATCC 68176 and ATCC 68175, respectively. The  
5 cre coding region can be isolated from plasmid Cre/Hpt-B with the restriction enzymes KpnI and SalI.

In one embodiment, the first, second, and optionally, third DNA sequences are introduced into one plant by transformation either in one step or in two or  
10 three successive steps. Alternatively, the first and second DNA sequences are introduced into one plant and the third DNA sequence into a different plant. The two plants are then sexually hybridized to produce progeny having all three DNA sequences. In another embodiment  
15 the first, second, and third DNA sequences are each introduced separately into a plant and the three are brought together by sexual hybridization.

Most preferably, the plasmid for introducing a DNA sequence comprising a promoter and a cre coding region  
20 is Cre/Hpt-A or Cre/Hpt-B and the plasmid for introducing a DNA sequence comprising a lox site is loxP/NptII/Hra or derivatives thereof carrying a pre-selected DNA segment other than or in addition to the Rubisco small subunit polyadenylation nucleotide  
25 sequence located between the first and second lox sites. These plasmids can be used to generate plants carrying cre or lox by those skilled in the art or as taught in this application. A Cre plant and a lox plant can be sexually hybridized to produce hybrid progeny plants  
30 containing a cre coding region and lox sites.

Since the lox site is an asymmetrical nucleotide sequence, the lox sites on the same DNA molecule can have the same or opposite orientation with respect to each other. Recombination between lox sites in the same  
35 orientation results in a deletion of the DNA segment

located between the two lox sites and a connection between the resulting ends of the original DNA molecule. The deleted DNA segment forms a circular molecule of DNA. The original DNA molecule and the resulting  
5 circular molecule each contain a single lox site. Recombination between lox sites in opposite orientations on the same DNA molecule result in an inversion of the nucleotide sequence of the DNA segment located between the two lox sites. In addition, reciprocal exchange of  
10 DNA segments proximate to lox sites located on two different DNA molecules can occur. All of these recombination events are catalyzed by the product of the cre coding region.

In a preferred embodiment of the present invention,  
15 the first and second DNA sequences are introduced into plant cells connected by a pre-selected DNA segment. The segment can be a gene or any other sequence of deoxyribonucleotides of homologous, heterologous or synthetic origin. Preferably, the pre-selected DNA  
20 segment is a gene for a structural protein, an enzyme, or a regulatory molecule; or a DNA sequence that influences gene expression in plant cells such as a regulatory nucleotide sequence, a promoter, or a polyadenylation nucleotide sequence. If the first and  
25 second lox sites have the same orientation, contact with Cre produces a deletion of the pre-selected DNA segment. If the first and second lox sites have opposite orientation, contact with Cre produces an inversion ("flipping") of the nucleotide sequence of the pre-  
30 selected DNA segment.

An effort was made to demonstrate the activation of gene expression using the flipping mode of the loxP-cre system. A construction was made in which the coding and polyadenylation regions from a sulfonylurea-resistant  
35 ALS gene were placed between two synthetic loxP sites

that were in inverted orientation relative to each other. This loxP bounded fragment was placed in inverted orientation to the 35S promoter such that it would not be expressed. The entire interrupted gene was  
5 put into a binary vector including a kanamycin resistance selection marker, introduced into A. tumefaciens, and then into tobacco plants. As expected, kanamycin-resistant transformants were not resistant to chlorsulfuron (a sulfonylurea),  
10 demonstrating no expression of the inverted coding region. Tissue was taken from selected transformants and retransformed using Agrobacterium containing -/Hpt or Cre/Hpt-B (Example 4). Hygromycin selected plants that received Cre retained their sensitivity to  
15 chlorsulfuron, indicating that the sulfonylurea-resistant ALS gene was not activated. The ALS gene was not activated because the loxP-bounded fragment did not flip in the plants. This was determined by analyzing plant DNA on Southern blots: a band representing the  
20 original lox construction was detected, but no band representing the flipped loxP-bounded fragment was detected. The loxP-bounded fragment was then shown to be incapable of flipping in an in vitro reaction using purified Cre. Thus this particular lox construction was  
25 defective in some as yet undetermined aspect. It is fully anticipated that if the loxP-bounded fragment were capable of flipping in the in vitro reaction, it would flip in plants containing Cre and the ALS gene would be activated.

30

#### UTILITY

The invention permits the site-specific recombination of DNA at the points of the introduced lox sites in any of the following ways:

- (a) Deletion of the DNA segment flanked by lox  
35 sites (excision);

(b) Inversion of the nucleotide sequence of the DNA segment flanked by lox sites (flipping);  
or

5 (c) reciprocal exchange of DNA segments proximate to lox sites located on different molecules (exchange).

Mode (a), excision, occurs when the lox sites are in like orientation on the same DNA molecule. One example of this event is to permit the removal of undesired  
10 marker genes, such as those that confer antibiotic resistance or herbicide resistance, in transgenic plants. Removal of the marker would also allow the use of the same marker in a second transformation of the transgenic plant. Also a trait gene that is undesired  
15 in a specific tissue or at a certain developmental time can be excised. Also a DNA sequence influencing expression of a gene can be excised resulting in increased or decreased expression of the gene. One skilled in the art will recognize that the reverse of  
20 excision (i.e., integration) may also be performed.

Mode (b), flipping, occurs when the lox sites are in reverse orientation on the same DNA molecule. This event may provide new methods of cre-regulated gene expression. Gene expression can be turned on by  
25 changing the direction of a promoter or regulatory nucleotide sequence from an inactive to an active orientation with respect to a coding region. Also changing the orientation of a coding region with respect to a promoter will alter its expression. Other ways to  
30 turn expression of a gene off include flipping an antisense RNA or ribozyme from an inactive to an active orientation.

Mode (c), exchange, may provide useful tools for recombinant alterations of plant DNA.



One application of the instant invention is in controlling male fertility in a method for producing hybrid crops. Hybridization of a crop involves the crossing of two different lines to produce hybrid seed from which the crop plants are grown. Hybrid crops are superior in that more of the desired traits can be introduced into the production plants. For instance, quality traits such as oil content, herbicide resistance, disease resistance, adaptability to environmental conditions, and the like, can be hybridized in offspring so that the latter are invested with the most desirable traits of its parents. In addition, progeny from a hybrid cross may possess new qualities resulting from the combination of the two parental types, such as yield enhancement resulting from the phenomenon known as heterosis. Controlled cross-fertilization to produce hybrid seeds has been difficult to achieve commercially due to competing self-fertilization, which occurs in most crop plants.

Currently, hybrid seed production is performed by one of the following means: (a) mechanically removing or covering the male organs to prevent self-fertilization followed by exposing the male-disabled plants to plants with male organs that contain the trait(s) desired for crossing; (b) growing genetically male-sterile plants in the presence of plants with fertile male organs that contain the trait that is desired for crossing; or (c) treating plants with chemical hybridizing agents (CHA) that selectively sterilize male organs followed by exposing the male-disabled plants to plants with fertile male organs that contain the trait that is desired for crossing. Some disadvantages to each of these methods include: (a) applicability only to a few crops, such as corn, where the male and female organs are well separated; and it is

labor intensive and costly; (b) genetically male sterile lines are cumbersome to maintain, requiring crosses with restorer lines; (c) all CHAs exhibit some degree of general phytotoxicity and female fertility reduction.

5 Also CHAs often show different degrees of effectiveness toward different crop species, or even toward different varieties within the same species.

10 A new molecular genetic approach to hybrid crop production that is applicable to a wide range of crops and involves genetic male sterility has been developed by Plant Genetic Systems. As described in EPA 89-344029, this system involves the introduction of a cell disruption gene that is expressed only in the tapetal tissue of anthers thereby destroying the  
15 developing pollen. The resulting genetically male sterile plants serve as the female parents in the cross to produce hybrid seed. This system could be highly effective and desirable. However one disadvantage is that since the male sterile parent is heterozygous for the sterility gene which acts as a dominant trait, only  
20 50% of the plants grown from the hybrid seed are fertile, the rest retain the sterility gene. This situation will result in reduced pollen shed in the production field which may lead to reduced seed set and yield. Addition of loxP-cre technology to this hybrid  
25 scheme will allow restoration of fertility to a much higher percentage of plants in the production field, as well as elimination of the cell disruption gene. Placing the male sterility gene between loxP sites  
30 allows it to be deleted following introduction of Cre into the hybrid from the male parent.

Another application of the instant invention is in making seedless produce. Seedlessness is desirable in consumed produce for convenience and taste. Currently  
35 "seedless" watermelon is sold that actually contains

some developed seed and a large number of immature seed that varies in size up to that of fully mature seed. To produce these watermelon first a hybrid cross is made between a tetraploid maternal parent and a diploid

5 pollinator. The resulting triploid seed produces self-infertile plants that are crossed with a diploid pollinator to produce seedless fruit [H. Kihara, Proc. Soc. Hort. Sci., 58:217-230, (1951)]. This production scheme suffers the following problems: (i) Creating a

10 tetraploid plant, which is accomplished by a chromosome duplication method, is difficult. Also the number of seeds per fruit on this tetraploid plant must be low since this has a positive correlation with seed number in the final product [C. F. Andrus, Production of

15 Seedless Watermelons, USDA Tech. Bull. No. 1425 (1971)]. (ii) Good combining ability of the diploid pollinator and the tetraploid plant is difficult to achieve [W. R. Henderson, J. Amer. Soc. Hort. Sci., 102:293-297 (1977)]. (iii) The triploid seeds are much inferior to

20 regular diploid seeds in vigor and germinability [D. N. Maynard, Hort. Sci., 24:603-604 (1989)]. These problems, together with incomplete seedlessness in the final product, make the development of seedless watermelon slow and difficult. This ploidy-based

25 approach to seedlessness is possible only in those few species where tetraploid and diploid plants are viable.

A molecular genetic approach to seedlessness involving loxP-cre is much more efficient, resulting in a more reliably seedless product and does not involve

30 changes in ploidy. Thus it is more generally applicable to a wider range of species. A lox/polyA-inactivated cell disruption gene regulated by a seed-specific promoter is introduced into a plant. When this plant is crossed to a plant expressing Cre, the disruption gene

35 is activated and expressed in the seed, thereby

disrupting seed development. The certainty of endosperm failure (caused by the cell disruption gene product) leading to the abortion of the whole seed is very high. In most dicots, the endosperm supplies the nutrients  
 5 needed for early embryo development. Endosperm abortion invariably leads to seed abortion [R. A. Brink and D. C. Cooper, Bot. Rev. 8: 423-541 (1947)].

The seed-specific promoter used is selected from the group of promoters known to direct expression in the  
 10 embryo and/or the endosperm of the developing seed, most desirably in the endosperm. Examples of seed-specific promoters include but are not limited to the promoters of seed storage proteins. The seed storage proteins are strictly regulated, being expressed almost exclusively  
 15 in seeds in a highly tissue-specific and stage-specific manner [Higgins et al., Ann. Rev. Plant Physiol. 35: 191-221 (1984); Goldberg et al., Cell 56: 149-160 (1989)]. Also, different seed storage proteins may be expressed at different stages of seed development and in  
 20 different parts of the seed.

There are numerous examples of seed-specific expression of seed storage protein genes in transgenic dicotyledonous plants. These include genes from dicotyledonous plants for bean  $\beta$ -phaseolin [Sengupta-  
 25 Goplalan et al., Proc. Natl. Acad. Sci. USA 82: 3320-3324 (1985) and Hoffman et al., Plant Mol. Biol. 11: 717-729 (1988)], bean lectin [Voelker et al., EMBO J 6: 3571-3577 (1987)], soybean lectin [Ocamuro et al., Proc. Natl. Acad. Sci. USA 83: 8240-8344 (1986)].  
 30 soybean kunitz trypsin inhibitor [Perez-Grau and Goldberg Plant Cell 1: 1095-1109 (1989)], soybean  $\beta$ -conglycinin [Beachy et al., EMBO J 4: 3047-3053 (1985), Barker et al., Proc. Natl. Acad. Sci. 85: 458-462 (1988), Chen et al., EMBO J 7: 297-302 (1988),  
 35 Chen et al., Dev. Genet. 10: 112-122 (1989), Naito

et al., Plant Mol. Biol. 11: 683-695 (1988)], pea vicilllin [Higgins et al., Plant Mol. Biol. 11: 109-123 (1988)], pea convicilllin (Newbiglin et al., Planta 180: 461 (1990)], pea legumin [Shirsat et al., Mol. Gen. Genetics 215: 326 (1989)], rapeseed napin [Radke et al., Theor. Appl. Genet. 75: 685-694 (1988)], as well as genes from monocotyledonous plants such as for maize 15-kd zein [Hoffman et al., EMBO J 6: 3213-3221 (1987)], barley  $\beta$ -hordein [Marris et al., Plant Mol. Biol. 10: 359-366 (1988)], and wheat glutenin [Colot et al., EMBO J 6: 3559-3564 (1987)]. Moreover, promoters of seed-specific genes operably linked to heterologous coding regions in chimeric gene constructions also maintain their temporal and spatial expression pattern in transgenic plants. Such examples include Arabidopsis thaliana 2S seed storage protein gene promoter to express enkephalin peptides in Arabidopsis and Brassica napus seeds [Vandekerckhove et al., Bio/Technology 7: 929-932 (1989)], bean lectin and bean  $\beta$ -phaseolin promoters to express luciferase [Riggs et al., Plant Sci. 63: 47-57 (1989)], and wheat glutenin promoters to express chloramphenicol acetyl transferase [Colot et al., EMBO J 6: 3559-3564 (1987)]. Promoters highly expressed early in endosperm development are most effective in this application. Of particular interest is the promoter from the  $\alpha'$  subunit of the soybean  $\beta$ -conglycinin gene [Walling et al., Proc. Natl. Acad. Sci. USA 83: 2123-2127 (1986)] which is expressed early in seed development in the endosperm and the embryo.

The cell disruption gene used is selected from a group of genes encoding products that disrupt normal functioning of cells. There are many proteins that are toxic to cells when expressed in an unnatural situation. Examples include the genes for the restriction enzyme EcoRI [Barnes and Rine, Proc. Natl. Acad. Sci. USA 82:

1354-1358 (1985)], diphtheria toxin A [Yamaizumi et al., Cell 15: 245-250 (1987)], streptavidin [Sano and Cantor, Proc. Natl. Acad. Sci. USA 87: 142-146 (1990)], and barnase [Paddon and Hartley, Gene 53: 11-19 (1987)].

- 5 Most preferred for this system is the coding region of barnase which has been shown to be highly effective in disrupting the function of plant cells (EPA 89-344029).

A highly desirable seedless system is one in which fully fertile F1 seed develops, that can then be grown  
 10 into plants that produce only seedless fruit. This system is economically favorable in that for each cross pollination, a large number of seedless fruits result: the number of F1 seed from one cross X the number of fruits produced on an F1 plant. Also incorporated in  
 15 this scheme are the advantages of growing a hybrid crop, including the combining of more valuable traits and hybrid vigor. This is accomplished in the same manner as described above except that the lox/polyA-inactivated disruption gene is expressed from a seed coat-specific  
 20 promoter. The seed coat is the outgrowth of the integuments, a strictly maternal tissue. Therefore the hybrid cross that brings the lox/polyA-inactivated disruption gene together with the cre gene does not involve this seed coat tissue. The seed coat of the F1  
 25 seed has either lox or cre, depending on which is used as the female parent, and thus F1 seed develop normally. After the F1 seed gives rise to a fruit-bearing F1 plant, all vegetative cells (including seed coat cells) inherit both lox and cre from the embryo. Thus the seed  
 30 coat of the F1 plant has an activated cell disruption gene.

The seed coat is an essential tissue for seed development and viability. When the seed is fully  
 35 matured, the seed coat serves as a protective layer to inner parts of the seed. During seed development, the

seed coat is a vital nutrient-importing tissue for the developing embryo. The seed is nutritionally "parasitic" to the mother plant. All raw materials necessary for seed growth must be imported. In seeds of dicotyledonous plants, the vascular tissue enters the seed through the funiculus and then anastomoses in the seed coat tissue. There is no vascular tissue connection or plasmodesmata linkage between the seed coat and the embryo. Therefore, all nutrient solutes delivered into the developing seed must be unloaded inside the seed coat and then move by diffusion to the embryo. Techniques have been developed to study the nutrient composition in the seed coat [Hsu et al., Plant Physiol. 75: 181 (1984); Thorne & Rainbird, Plant Physiol. 72: 268 (1983); Patrick, J. Plant Physiol. 115: 297 (1984); Wolswinkel & Ammerlaan, J. Exp. Bot. 36: 359 (1985)], and also the detailed cellular mechanisms of solute unloading [Offler & Patrick, Aust. J. Plant Physiol. 11: 79 (1984); Patrick, Plant Physiol. 78: 298 (1990)]. It is obvious that the destruction of this vital nutrient-funnelling tissue causes seed abortion.

#### EXAMPLES

##### SITE SPECIFIC RECOMBINATION IN PLANTS

##### Materials and Methods

##### 25 Molecular Techniques

Methods of culturing bacteria, preparing DNA, and manipulating DNA were as described by Maniatis et al., Molecular Cloning: A Laboratory Manual [Cold Spring Harbor Laboratory, New York (1982)] unless stated otherwise. Restriction enzymes and other enzymes used in DNA manipulations were obtained from New England Biolabs, Inc. (Beverly, MA, USA), Boehringer Mannheim (Indianapolis, IN, USA), or Bethesda Research Laboratory (Gaithersburg, MD, USA) and were used essentially according to the manufacturer's specifications.

### Isolation and analysis of plant RNA and DNA

Both RNA and DNA were extracted from the same leaf samples by combining methods for extraction of each. One to five grams of leaf tissue were frozen in liquid nitrogen and ground. Frozen tissue was added to 15 ml of extraction buffer [100 mM Trizma hydrochloride (Tris) pH 8.0, 50 mM EDTA pH 8.0, 100 mM NaCl, 1% sodium dodecyl sulfate (SDS), 200 µg/ml proteinase K] and heated at 65°C for 10 min. Five ml of 5 M potassium acetate was added, and the samples were placed on ice for 20 min. The samples were spun at 25K x g for 20 min and the supernatant was poured through cheesecloth into a tube containing 1 ml of 5 M sodium acetate and 10 ml of isopropanol. The tubes were left overnight at -20°C. The RNA/DNA was pelleted by centrifugation at 20K x g for 15 min. The pellets were resuspended in 10 mls of water and an equal volume of 4 M lithium chloride was added. The solutions were placed on ice for 1-2 hours, then centrifuged for 20 min at 20K x g. The supernatant was collected and an equal volume of isopropanol was added. After an overnight incubation at -20°C, the DNA was pelleted and resuspended in a solution of 10 mM Tris with 1 mM EDTA pH 8.0 (TE). The samples were extracted with an equal volume of Tris pH 8.0 buffered phenol and precipitated by adding 0.1 volume of 3 M sodium acetate and 2.5 volumes of ethanol. For Southern blot analysis the DNA was digested with a restriction enzyme and the resulting fragments were separated by gel electrophoresis, transferred to Zeta-Probe filters (Bio-Rad Laboratories, Richmond, CA, USA), and hybridized with nick translated probes.

The lithium chloride pellet was resuspended in one-half the original volume of water, an equal volume of lithium chloride was added, and the mixture was placed on ice for an additional hour. The RNA was pelleted by



centrifugation, resuspended in water, extracted with buffered phenol, and precipitated with 0.1 volume of 3 M sodium acetate and two volumes ethanol. For Northern blot analysis the RNA was separated by gel

- 5 electrophoresis in formaldehyde as described by Rave et al., Nucl. Acids Res. 6: 3559-3569 (1979), transferred to Zeta-Probe filters, and hybridized to nick translated probes.

#### Generation of transgenic plants

- 10 The cointegrate and binary Ti plasmids containing the chimeric cre gene and those containing loxP sites were introduced into tobacco by leaf disk transformation and into Arabidopsis by root transformation as described in Example 8. Standard aseptic techniques for the
- 15 manipulation of sterile media and axenic plant/bacterial cultures were followed, including the use of a laminar flow hood for all transfers. Recipes for media for tobacco are given in Table 2. Potted tobacco plants for leaf disk infections were grown in a growth chamber
- 20 maintained for a 14 hr, 24°C day, 10 hr, 20°C night cycle, with approximately 80% relative humidity, under mixed cool white fluorescent and incandescent lights. Tobacco leaf disk infections were carried out essentially by the method of Horsch et al., Science 227,
- 25 1229-1231 (1985), omitting nurse cultures.

- Healthy young leaves, not fully expanded and approximately 3-5 inches in length, were harvested from approximately 4-6 week old tobacco plants (Nicotiana glauca var. Xanthi). The leaves were surface
- 30 sterilized by immersion in a solution containing 10% commercial bleach and 0.1% sodium dodecyl sulfate (SDS). After 20 minutes of sterilization with intermittent mixing, the leaves were transferred successively three times to sterile deionized water to rinse the leaves
- 35 thoroughly, and then shaken gently to remove excess

water. Leaf disks, 8 mm in diameter, were prepared from whole leaves using a sterile paper punch.

Cultures of Agrobacterium cells containing the binary or cointegrate plasmids were grown in 5 ml of YEB or YEP broth (Table 1 and Table 8) with the appropriate antibiotics. Cultures were grown for approximately 17-20 hours in 18 mm glass culture tubes in a New Brunswick platform shaker maintained at 28°C. Leaf disks were inoculated by submerging them for several minutes in 20 ml of a 1:20 dilution of the overnight Agrobacterium culture.

After inoculation, the leaf disks were placed in petri dishes containing .1N1B agar medium (Table 2). The dishes were sealed with parafilm and incubated under mixed fluorescent and "Gro and Sho" plant lights (General Electric) for 2-3 days in a culture room maintained at approximately 25°C.

To rid the leaf disks of Agrobacterium and to select for the growth of transformed tobacco cells, the leaf disks were transferred to fresh .1N1B medium containing 500 mg/l cefotaxime and either 10-30 mg/l hygromycin, most preferably 30 mg/l hygromycin, 20-50 ppb chlorsulfuron, most preferably 25 ppb chlorsulfuron, or 100-300 mg/l kanamycin, most preferably 100 mg/l kanamycin. Cefotaxime was kept as a frozen 200 mg/ml stock solution and added aseptically (filter sterilized through a 0.45  $\mu$ m filter) to the media after autoclaving. A fresh stock of hygromycin, chlorsulfuron, or kanamycin was made for each use and was filter sterilized into the autoclaved media.

Leaf disks were incubated under the growth condition described above for 2-3 weeks and then transferred to fresh media of the same composition or to MX<sup>-</sup> with the appropriate antibiotics.

- Approximately 3-4 weeks later, shoots developing on medium containing either hygromycin, chlorsulfuron, or kanamycin were excised with a sterile scalpel and planted in MX<sup>-</sup> medium (Table 2) containing 200-500 mg/l cefotaxime, most preferably 250 mg/l cefotaxime, in the presence or absence of 10-30 mg/l hygromycin, most preferably 30 mg/l hygromycin, or 20-30 ppb chlorsulfuron, most preferably 25 ppb chlorsulfuron. Root formation was recorded within 3 weeks.
- Leaves were removed from the rooted excised shoots to determine levels of resistance to hygromycin, chlorsulfuron, or kanamycin in a callus induction assay on selective media. To induce callus formation, leaves were excised and leaf disks, 8 mm in diameter, were made using a sterile paper punch and plated on callus induction medium (Table 2) containing either 20-50 mg/l hygromycin, most preferably 30 mg/l hygromycin, 5-25 ppb chlorsulfuron, most preferably 25 ppb chlorsulfuron, or 50-100 mg/l kanamycin, most preferably 100 mg/l kanamycin. Callus growth on selective and non-selective media was recorded within 3 weeks.

TABLE 1

AGROBACTERIUM GROWTH MEDIA

## 25 YEB MEDIUM

	<u>Per Liter</u>
Bacto Beef Extract	5.0 g
Bacto Yeast Extract	1.0 g
Peptone	5.0 g
30 Sucrose	5.0 g
MgSO <sub>4</sub> ·7H <sub>2</sub> O	0.5 g
Agar (optional)	15.0 g
pH 7.2	

## MIN A with sucrose plates

		<u>Per Liter</u>
	water	948 ml
	agar	15 g
5	mix and autoclave	
	min A salts:	40 ml
	200 ml: $K_2HPO_4$	52.5 g
	$KH_2PO_4$	22.5 g
	$(NH_4)_2SO_4$	5 g
10	Na Citrate $2H_2O$	2.5 g
	20% $MgSO_4 \cdot 7H_2O$	1 ml
	1% thiamine hydrochloride	0.5 ml
	20% sucrose	10 ml

15

TABLE 2TOBACCO TISSUE CULTURE MEDIACallus Induction Medium

		<u>Per Liter</u>
20	Murashige's Minimal Organics Medium	1 package
	GIBCO #510-1118 (contains 3% sucrose)	
	100X Vitamin Supplement:	10 ml
	10 mg/l thiamine	
	50 mg/l pyridoxine	
25	50 mg/l nicotinic acid	
	1 mg/ml naphthaleneacetic acid	1 ml
	(NAA) stock	
	1 mg/ml 6-Benzylaminopurine (BAP) stock	0.2 ml
	agar	8.0 g
30	pH 5.8	

Shoot Induction Medium (.1N1B)

	<u>Per Liter</u>
Murashige's Minimal Organics Medium	1 package
GIBCO #510-1118 (contains 3% sucrose)	
100X Vitamin Supplement:	10 ml
10 mg/l thiamine	
50 mg/l pyridoxine	
50 mg/l nicotinic acid	
1 mg/ml NAA (naphthaleneacetic acid) stock	0.1 ml
1 mg/ml BAP stock	1.0 ml
agar	8.0 g
pH 5.8	

Root Induction Medium (MX<sup>-</sup>)

	<u>Per Liter</u>
Murashige's Minimal Organics Medium	1 package
GIBCO #510-1118 (contains 3% sucrose)	
100X Vitamin Supplement:	10 ml
10 mg/l thiamine	
50 mg/l pyridoxine	
50 mg/l nicotinic acid	
agar	8.0 g
pH 5.8	

A. Construction of plasmids for integration  
and expression of the cre coding region  
in plant cells

EXAMPLE 1

The starting material for construction of Cre/Hpt-A was the plasmid pK35CAT, which is described in Lin et al., Plant Physiology, 84: 856-861 (1987) and has been deposited in the ATCC and bears deposit accession number 68174. This plasmid contains a CaMV 35S promoter (35S/P) directing expression of the chloramphenicol

acetyltransferase (CAT) coding region and followed by a nopaline synthase (NOS) gene polyadenylation nucleotide sequence (NOS 3'). pK35CAT was derived from pK35K, which was in turn derived from pKNK. pKNK has been deposited with the ATCC and bears the deposit accession number 67284. pKNK is a pBR322 based vector which contains a neomycin phosphotransferase II (NptII) promoter fragment, a nopaline synthase (NOS) promoter fragment, the coding region of NptII and the polyadenylation region from the NOS gene. A map of this plasmid is shown in Lin et al., Plant Physiol. 84: 856-861 (1987). The 320 bp ClaI-BglIII fragment in pKNK that contains the NptII promoter was obtained as a HindIII-BglIII fragment from the NptII gene of the transposon Tn5 described by Beck et al., Gene 19: 327-336 (1982). The HindIII site was converted to a ClaI site by linker addition. The NptII promoter fragment is followed by a 296 bp Sau3A-PstI NOS promoter (NOS/P) fragment corresponding to nucleotides -263 to +33, with respect to the transcription start site, of the NOS gene described by Depicker et al., J. Appl. Genet. 1: 561-574 (1982). The PstI site at the 3' end was created at the translation initiation codon of the NOS gene. The NOS/P is followed by a 998 bp HindIII-BamHI sequence containing the NptII coding region obtained from the transposon Tn5 [Beck et al., Gene 19: 327-336 (1982)] by the creation of HindIII and BamHI sites at nucleotides 1540 and 2518, respectively. The NptII coding region is then followed by a 702 bp BamHI-ClaI fragment containing the 3' end of the nopaline synthase gene including nucleotides 848 to 1550 [Depicker et al., J. Appl. Genet. 1: 561-574 (1982)]. The remainder of pKNK consists of pBR322 sequences from 29 to 4361.

pKNK was converted to pK35K by removing the NptII and NOS promoters and replacing them with a CaMV 35S promoter. The EcoRI-HindIII 35S promoter fragment is the same as that contained in pUC35K which has been deposited with the ATCC and bears the deposit accession number 67285. The 35S promoter fragment was prepared as follows, and as described in Odell et al., Nature 313: 810-813 (1985) except that the 3' end of the fragment includes CaMV sequences to +21 with respect to the transcription start site. A 1.15 kb BglII segment of the CaMV genome containing the region between -941 and +208 relative to the 35S transcription start site was cloned in the BamHI site of the plasmid pUC13. This plasmid was linearized at the SalI site in the polylinker located 3' to the CaMV fragment and the 3' end of the fragment was shortened by digestion with nuclease Bal31. Following the addition of HindIII linkers, the plasmid DNA was recircularized. From nucleotide sequence analysis of the isolated clones, a 3' deletion fragment was selected with the HindIII linker positioned at +21. To create pK35K this 35S promoter fragment was isolated as an EcoRI-HindIII fragment, the EcoRI site coming from the polylinker of pUC13, and ligated to pKNK that had been digested with EcoRI and HindIII, the EcoRI site lying 5' to the ClaI site in pBR322.

pK35K was converted to pK35CAT by dropping out the NptII coding region and replacing it with the coding region of chloramphenicol acetyl transferase (CAT) as mapped and described in Lin et al., Plant Physiol. 84: 856-861 (1987). The CAT coding region was obtained as a 975 bp Sau3A fragment from pBR325. The ends were filled in and the fragment was ligated into a filled in SalI site of pGEM2. A clone, pGCAT9, was selected that contains the insert oriented such that the HindIII and

BamHI sites of the polylinker are located 5' and 3' to the CAT coding region, respectively. The CAT coding region was isolated from this clone by HindIII and BamHI digestion, and ligated into HindIII and BamHI digested pK35K. The resultant construction, termed pK35CAT, also contains the NOS 3' fragment which remains unaltered in the conversion of pKNK to pK35K, and finally to pK35CAT.

The entire cre gene was originally obtained from the genome of bacteriophage P1 on an EcoRI fragment as described by Sternberg and Hamilton, J. Mol. Biol. 150: 467-486 (1981). The cre coding region was prepared as an XhoI-EcoRI fragment in plasmid pRH103Δ6 as described by Sternberg et al., J. Mol. Biol. 187: 197-212 (1986). The XhoI site was added as a linker following Bal31 deletion of the sequence 5' to the cre coding region, resulting in the placement of the XhoI site approximately 50 bp 5' to the translation initiation ATG. The EcoRI site was added as a linker following Bal31 deletion of the sequence 3' to the cre coding region, resulting in the placement of the EcoRI site approximately 100 bp 3' to the translation stop codon. The 3' EcoRI site was then replaced with a SalI site generating pBS7 as described in Sauer, Mol. and Cell. Biol. 7: 2087-2096 (1987) so that the cre coding region could be isolated as a XhoI-SalI fragment. This cre coding fragment is the same as that present in plasmid pBS39 which has been deposited with the ATCC and bears deposit accession number 53255. The XhoI-SalI cre coding region fragment was isolated, HindIII linkers were added to the ends, and it was ligated with HindIII digested pK35CAT, generating pK35CreCAT. This plasmid contains a chimeric 35S/P-cre-CAT-NOS 3' gene.

To construct Cre/Hpt-A, pK35CAT was digested with BamHI, the end was partially filled with dGTP and dATP according to the method of Hung and Wensink, Nucl. Acids



Res. 12: 1863-1874 (1984) and then it was digested with HindIII to remove the CAT coding region. The HindIII-SalI DNA fragment containing the cre coding region was isolated from pK35CreCAT, whose construction is  
 5 described in the previous paragraph, and the SalI site was partially filled with dCTP and dTTP during its preparation. This cre coding region fragment was then ligated into the prepared vector derived from pK35CAT creating the plasmid pK35Cre which contains a chimeric  
 10 35S promoter-cre coding region-NOS 3' gene.

#### EXAMPLE 2

Next a ClaI-SalI fragment containing a chimeric NOS/P-Hpt-NOS 3' gene (Hpt=hygromycin phosphotransferase) and the NptI gene (neomycin  
 15 phosphotransferase I) was isolated from pAGS122 which is analogous to pAGS120 that is described in van den Elzen et al., Plant Molecular Biology, 5: 299-302 (1985). A SalI linker was added to the ClaI end of the fragment and it was ligated into SalI digested pK35Cre creating  
 20 the plasmid Cre/Hpt-A which is shown in Figure 1A.

The boxes represent the chimeric 35S/P-cre-NOS 3' and NOS/P-Hpt-NOS 3' chimeric genes. The arrows represent the transcripts expressed by these chimeric genes. The NptI gene is derived from Tn903. These  
 25 genes are incorporated in a pBR322 vector.

#### EXAMPLE 3

The starting material for construction of Cre/Hpt-B was the plasmid pDH51 that was described by Pietrzak et al., Nucleic Acids Research, 14: 5857-5868 (1986). This  
 30 plasmid contains a CaMV 35S promoter including sequences between 6909 and 7437 of the CaMV genome and a CaMV polyadenylation nucleotide sequence including sequences between 7439 and 7632, separated by several restriction enzyme sites, including XbaI. The CaMV promoter  
 35 fragment in pDH51 was prepared by adding an EcoRI linker

5' to the NcoI site at 6909 of the CaMV genome and a KpnI linker at the HphI site at 7437. The polyadenylation region fragment was prepared by adding an SphI linker at the HphI site at 7439 and a HindIII linker following KpnI, SstI, and EcoRI sites that had been added onto position 7632 during a cloning step in pUC18. Both of these fragments were cloned into pUC18 using the restriction sites located on their ends to generate pDH51. In the resulting plasmid, EcoRI sites are located on either end outside of the CaMV promoter and 3' region. pDH51 was digested with XbaI and the ends were partially filled with dCTP and dTTP. The HindIII DNA fragment containing the cre coding region was isolated from pk35CreCAT and the ends were partially filled with dATP and dGTP, then ligated into the prepared pDH51 vector. To identify a plasmid with the HindIII cre fragment in the proper orientation for expression, a SalI digest was done. The desired plasmid was digested with SalI since the SalI site at the 3' end of the cre fragment is adjacent to the SalI site at the 5' end of the CaMV 3' region. A BamHI digest confirmed the correct orientation: a 430 bp fragment between the BamHI site at the 3' end of the 35S/P region and a BamHI site within the cre coding region was present. The resulting pDH51Cre plasmid contains a chimeric 35S promoter-cre coding region-CaMV 3' gene.

#### EXAMPLE 4

Next the entire chimeric gene was cloned into pJJ2644, which is a binary vector for Agrobacterium tumefaciens transformation that carries a chimeric 1'/P-Hpt-NOS 3' gene, a tetracycline resistance gene, a broad host range origin of replication, and T-DNA borders. pJJ2644 has been deposited with the ATCC and bears deposit accession number 68178 and was constructed as follows. The broad host range plasmid pRK290 that

was described by Ditta et al., Proc. Natl. Acad. Sci. USA 77: 7347-7351 (1980) served as the basic vector.

This plasmid was cut with EcoRI, the ends filled, and it was ligated to an end site filled EcoRI-HindIII fragment

5 isolated from pAGS111 creating pJJ1881. The fragment from pAGS111 contains the left and right border fragments from the Agrobacterium tumefaciens T-DNA located on either side of a chimeric NptII gene and its construction is described in van den Elzen et al., Plant Molec. Biol. 5: 149-154 (1985). The ClaI-BamHI chimeric NptII gene fragment was replaced with the ClaI-BamHI fragment from pBR322, thereby adding a HindIII site. To create pJJ2501 a HindIII-ClaI fragment was added that contains a chimeric 1'/P-Hpt-NOS 3' gene consisting of  
15 the 1' promoter described by Velten et al., EMBO J. 12: 2723-2730 (1984), the Hpt coding region described by van den Elzen et al., Plant Molec. Biol., 5: 299-302 (1985), from which the ATG sequence located just 5' to the translation initiation ATG had been removed, and the  
20 NOS 3' region. Next the XhoI site located outside of the T-DNA borders was deleted. Between the BamHI and HpaI sites, located 3' to the chimeric Hpt gene, a linker, including sites for BamHI, XbaI, HindIII, XhoI, EcoRI, and HpaI, was added creating pJJ2644.

25 This vector was digested with HindIII and ligated to the HindIII linkered EcoRI fragment from pDH51Cre containing the chimeric cre gene. The resulting plasmid is Cre/Hpt-B which is shown in Figure 1B. The boxes represent the chimeric 35S/P-cre-CaMV 3' and 1'/P-Hpt-NOS 3' genes. The T-DNA left and right borders are  
30 marked as filled boxes. The slashes indicate unrepresented sequences of the binary vector PJJ2644, which includes a tetracycline resistance gene (tet).

A plasmid used as a control is called -/Hpt-B and is  
35 the pJJ2644 vector with no chimeric cre gene added.

B. Construction of plasmids containing the lox site  
for integration and analysis in plant cells

EXAMPLE 5

5       The starting material for construction of  
loxP/NptII/Hra was the plasmid pBS69 containing two loxP  
sites which is described by Sauer and Henderson, Nucleic  
10 Acids Research, 17: 147-161 (1989). The loxP site was  
originally obtained from the bacteriophage P1 genome on  
a BamHI fragment cloned in pBR322 as described by  
Abremski et al., Cell 32: 1301-1311 (1983). This  
reference also describes construction of an 80 bp  
EcoRI-HindIII fragment containing loxP made by adding an  
EcoRI linker to the BclI site and a HindIII linker to  
15 the PvuII site, located on either side of loxP. A 50 bp  
BamHI-XhoI fragment containing loxP was made by deleting  
in from the BclI site with Bal31 and adding a BamHI  
linker 10 bp from loxP, and deleting in from the PvuII  
site and adding an XhoI linker 6 bp from loxP. The 80  
20 bp EcoRI-HindIII fragment containing loxP was ligated  
between the EcoRI and HindIII sites in the pBR322  
plasmid containing the 50 bp loxP fragment resulting in  
the plasmid pRH42 that has two loxP sites oriented in  
the same direction. A derivative of this plasmid called  
25 pRH43 having the NptII gene from Tn5 between the loxP  
sites is also described. A derivative of pRH43 called  
pRH499 that has the Leu2 gene of yeast between the loxP  
sites is described and a map shown in Sauer, Mol. and  
Cell. Biol. 7: 2087-2096 (1987). As described in this  
30 reference, the HindIII site adjacent to the 80 bp lox  
site was deleted generating pBS30. pBS30 carries the  
same EcoRI-XhoI fragment containing two directly  
oriented loxP sites that is present in pBS44 which has  
been deposited with the ATCC and bears deposit accession  
35 number 53254. As described in Sauer and Henderson

[Nucl. Acids Res. 17: 147-161 (1989)] pBS69 was generated from pBS30 by replacing the 80 bp HindIII-SalI fragment containing the loxP site with the 50 bp HindIII-XhoI fragment containing the loxP site. This was done to remove extra sequence in the 80 bp fragment that contained ATG translation initiation codons. Thus pBS69 has two directly oriented 50 bp loxP containing fragments surrounding a yeast Leu2 gene.

Part of the Leu2 gene was removed using the EcoRI site located within the Leu2 gene and the BamHI site adjacent to one loxP site, and replaced with a polyadenylation nucleotide sequence (polyA) derived from the tobacco Rubisco small subunit gene described by Mazur and Chui [Nucleic Acids Research, 13: 2373-2386 (1985)]. This reference describes the cloning and sequencing of this gene. The BamHI-XbaI fragment containing the sequence region between 1905 and 2289 was isolated, the XbaI site being filled in during its preparation. pBS69 was digested with EcoRI, the end filled in, then digested with BamHI. The ligation of these two DNAs resulted in the plasmid pBS69polyA.

#### EXAMPLE 6

Next an XhoI-HindIII fragment containing the loxP-polyA-loxP region was isolated and a HindIII linker was added to the XhoI end. This fragment was ligated into HindIII digested pKNK. The construction of pKNK was described in Example 1 and it contains the NOS promoter joined to the NptII coding region by a HindIII site resulting in a chimeric gene which confers kanamycin resistance to plant cells. The orientation of the loxP-polyA-loxP HindIII insert was determined by digestion with BamHI. The plasmid with the HindIII fragment inserted such that the polyA site has the same orientation as the NOS/P and the NptII coding region was

called pKNKloxA. It was anticipated that the polyadenylation nucleotide sequence located between the NOS/P and NptII coding region would block production of a viable NptII transcript thereby causing transformed  
 5 cells to retain their kanamycin sensitivity.

Next a PstI fragment that contains the Hra (sulfonylurea resistant acetolactate synthase) gene derived from pALS032BV that is described by Lee et al. [EMBO J. 7: 1241-1248 (1988)], and also contains a  
 10 HindIII fragment with a streptomycin/spectinomycin resistance gene derived from the R100.1 plasmid and described by Prentki and Krisch, *Gene*, 29: 303-313 (1984) was added. This fragment was constructed by adding SalI linkers to the HindIII ends of the isolated  
 15 strep/spec fragment and ligating it into the SalI site adjacent to the Hra gene in pALS032BV. The PstI fragment was then isolated, the ends filled in, and it was ligated into SalI digested and filled pKNKloxA. The resulting plasmid called loxP/NptII/Hra is shown in  
 20 Figure 1C.

In Figure 1C, the open boxes represent the chimeric NOS/P-NptII-NOS 3' gene that is interrupted between the promoter and coding region by a Rubisco small subunit gene polyadenylation region, shown as the stipled box,  
 25 which is surrounded by two loxP sites, represented by arrows showing that the loxP sites are in the same orientation. The asterisk marks the polyadenylation site. The plasmid includes the sulfonylurea-resistant ALS gene called Hra and the streptomycin and  
 30 spectinomycin resistance marker, incorporated in a pBR322 vector. The orientation of the PstI insert was not determined.

C. Transformation of tobacco with the  
cre coding region

The chimeric 35S/P-Cre-NOS 3' gene, described in Examples 1 and 2, was introduced into tobacco by  
5 Agrobacterium tumefaciens infection of tobacco leaf disks. Primary transformants were analyzed to demonstrate the presence of the cre coding region and expression of the cre mRNA in tobacco cells as well as expression of the linked Hpt gene which confers  
10 hygromycin resistance.

The plasmid Cre/Hpt-A was transferred into A. tumefaciens by a method involving a three-way mating that was essentially as described by Fraley et al. [Proc. Natl. Acad. Sci. USA, 80: 4803-4807 (1983)]  
15 except for the following points. Cre/Hpt-A was mated into Agrobacterium strain GV3850 that was described by Zambryski et al. [J. of Mol. and Appl. Genetics, 1: 361-370 (1982)]. Colonies from the Cre/Hpt-A mating were selected on LB plates containing 100 µg/ml rifampicin  
20 and 25 µg/ml kanamycin. Selected colonies were confirmed as cointegrates of the Cre/Hpt-A plasmid into the Ti plasmid by Southern blot analyses.

Standard aseptic techniques for the manipulation of sterile media and axenic plant/bacterial cultures were  
25 followed, including the use of a laminar flow hood for all transfers. Potted tobacco plants for leaf disk infections were grown in a growth chamber maintained for a 14 hr, 24°C day, 10 hr, 20°C night cycle, with approximately 80% relative humidity, under mixed cool  
30 white fluorescent and incandescent lights. Tobacco leaf disk infection was carried out essentially by the method of Horsch et al. [Science 227, 1229-1231 (1985)], omitting nurse cultures, as described below.

Healthy young tobacco leaves were harvested, surface  
35 sterilized, and rinsed as described in Materials and

Methods. Leaf disks, 8 mm in diameter, were prepared from whole leaves using a sterile paper punch.

Leaf disks were inoculated by submerging them for several minutes in 20 ml of a 1:20 dilution of the overnight culture of Agrobacterium harboring the cointegrate Cre/Hpt-A Ti plasmid. The culture was started by inoculating 5 mls of YEB broth (Table 1) with a single bacterial colony. The culture was grown for approximately 17-20 hours in 18 mm glass culture tubes in a New Brunswick platform shaker maintained at 28°C.

After inoculation, the leaf disks were placed on .1N1B agar medium (Table 2) in petri dishes which were then sealed with parafilm. The petri dishes were incubated under mixed fluorescent and "Gro and Sho" plant lights (General Electric) for 2-3 days in a culture room maintained at approximately 25°C.

To rid the leaf disks of Agrobacterium and to select for the growth of transformed tobacco cells, the leaf disks were transferred to fresh .1N1B medium containing 500 mg/L cefotaxime and 10-20 mg/l hygromycin. Cefotaxime was kept as a frozen 200 mg/ml stock solution and added aseptically (filter sterilized through a 0.45 µm filter) to the media after autoclaving. A fresh hygromycin stock (20 mg/ml) was made for each use and was filter sterilized into the autoclaved media. Leaf disks were incubated under the growth conditions described above for 3 weeks and then transferred to fresh media of the same composition.

Approximately 1 month later, shoots developing on hygromycin-containing medium were excised with a sterile scalpel and planted in MX<sup>-</sup> medium containing 200 mg/L cefotaxime and 10 mg/L hygromycin. Root formation was recorded within 3 weeks.

Leaves were removed from the rooted excised shoots to determine levels of resistance to hygromycin in a



callus induction assay on selective media. To induce callus formation, leaves were excised and leaf disks, 8 mm in diameter, were made using a sterile paper punch and planted on callus induction medium containing 20 and  
5 50 mg/l hygromycin. Callus growth on selective and non-selective media was recorded within 3 weeks.

The results shown in Table 3 indicate that transformation of tobacco had been achieved with Agrobacterium harboring the Cre/Hpt-A Ti plasmid, based  
10 on production of hygromycin resistant callus. All ten transformants tested were resistant to hygromycin. Primary transformants were analyzed by molecular techniques to verify the presence of the cre mRNA sequences. Seven independent tobacco plants transformed  
15 with Cre/Hpt-A were assayed for expression of the chimeric cre gene by Northern blots as described in Materials and Methods. The probe used to hybridize to the filter containing the RNA prepared from each plant was a BamHI-ClaI DNA fragment that was isolated from the  
20 pK35K or pKNK plasmid, which are both described in Example 1. This fragment contains the NOS polyadenylation nucleotide sequence which includes a region of untranslated transcribed sequence. Since both the 35S/P-Cre-NOS 3' and NOS/P-Hpt-NOS 3' genes have  
25 homology to this probe, the transcript from each gene is detected in this experiment. The expected 2.0 kb transcript from the cre gene and the expected 1.5 kb transcript from the Hpt gene were detected on the Northern filter. All of the plants assayed, except for  
30 one, produced a detectable level of Cre transcript indicating the presence and expression of the chimeric cre gene in the plant cells.

Plants exhibiting hygromycin resistance were transferred to soil and grown to maturity in a growth  
35 chamber as described above. Individual inflorescences

were covered with bags to permit self-fertilization without cross-pollination. Mature seeds were harvested and progeny tests were conducted to determine the inheritance of the introduced Cre/Hpt DNA. Inheritance was monitored by following the hygromycin resistance trait.

Seed was surface sterilized for 30 minutes in 10% commercial bleach and 1% SDS with intermittent mixing, rinsed 3-5 times with sterile deionized water, dried, and planted on MX<sup>-</sup> medium in the presence or absence of 50 mg/l hygromycin. Sensitive seeds germinated, but did not develop further. A segregation ratio of 3 resistant progeny to 1 sensitive indicated the presence of a single site of integration of the hygromycin resistance gene into the genome of the transformant, which was then stably inherited by its progeny. This was seen in seven out of eight independent transformants tested. The eighth transformant exhibited a ratio which was greater than 3:1, indicating the presence of more than one integration site.

TABLE 3  
Callus Growth on Hygromycin

ID	WEIGHT IN GRAMS	Hygromycin Concentration		
		0 mg/l	20 mg/l	50 mg/l
5	U1	9.02	4.55	3.36
	U2a	8.88	2.34	2.43
	U2b	7.72	3.77	2.88
	U2c	3.69	1.32	0.85
	C1= U3a <sup>1</sup>	4.43	2.42	1.28
10	C2= U3b <sup>1</sup>	8.32	3.63	2.02
	U4a <sup>1</sup>	6.21	2.52	1.15
	C3= U4b	5.80	2.94	2.82
	U5	4.23	3.28	2.49
	U6a <sup>1</sup>	5.79	2.16	2.00
15	C4= U6b	5.52	2.40	1.42
	U6c	11.37	2.56	2.10
	C5= U7a	8.41	2.78	2.16
	C6= U7b	5.22	4.66	4.23
	U7c	8.52	3.38	2.88
20	U8	4.38	5.15	2.09
	WT	3.55	0.68	0.36

<sup>1</sup> Weight in grams is the average of results from two callus induction assays done with the same transformant.

25                    D. Production of plant cells and plants  
                      with two loxP sites integrated into the genome

                      The loxP-polyA-loxP DNA sequence described in  
Examples 5 and 6 was introduced into tobacco by  
Agrobacterium tumefaciens infection of tobacco leaf  
30 disks as described in Materials and Methods. Primary  
transformants were analyzed to demonstrate the presence  
of the loxP-polyA DNA sequence in the tobacco cells as

well as expression of the linked Hra gene which confers resistance to chlorsulfuron.

The plasmid loxP/NptII/Hra was transferred into A. tumefaciens by a method involving a three-way mating that was essentially as described by Fraley et al. [Proc. Natl. Acad. Sci. USA, 80, 4803-4807 (1983)] except for the following points. LoxP/NptII/Hra was mated into Agrobacterium strain GV3850 that was described by Zambryski et al. [J. of Mol. and Appl. Genetics, 1: 361-370 (1982)]. Colonies from the loxP/NptII/Hra mating were selected on 100 µg/ml rifampicin and 100 µg/ml each of spectinomycin and streptomycin. Selected colonies were confirmed as cointegrates of the loxP/NptII/Hra plasmid into the Ti plasmid by Southern blot analyses.

Tobacco leaf disks were obtained, inoculated with A. tumefaciens harboring the cointegrate loxP/NptII/Hra plasmid, and incubated on .1N1B as described in Materials and Methods.

To rid the leaf disks of Agrobacterium and to select for the growth of transformed tobacco cells, the leaf disks were transferred to fresh .1N1B medium containing 500 mg/l cefotaxime and 20-200 ppb chlorsulfuron. Cefotaxime was kept as a frozen 200 mg/ml stock solution and added aseptically (filter sterilized through a 0.45 µm filter) to the media after autoclaving. A fresh chlorsulfuron stock was prepared for each use by first making a 0.2 mg/ml solution in 0.01 N NH<sub>4</sub>OH, which was then diluted 1:10 with deionized water, and filter sterilized into the autoclaved media. Leaf disks were incubated under the growth conditions described above for 3 weeks and then transferred to fresh media of the same composition.

Approximately 1 month later, shoots developing on medium containing 20-50 ppb chlorsulfuron were excised

with a sterile scalpel and planted in MX<sup>-</sup> medium containing 200 mg/l cefotaxime and 20 mg/l chlorsulfuron. Root formation was recorded within 3 weeks.

- 5        Leaves were removed from the rooted excised shoots to determine levels of resistance to chlorsulfuron and kanamycin in a callus induction assay on selective media. To induce callus formation, leaves were excised and leaf disks, 8 mm in diameter, were made using a  
10       sterile paper punch and planted on callus induction medium containing 5, 10, or 20 ppb chlorsulfuron and on callus induction medium containing 100 mg/l kanamycin. Callus growth on selective and non-selective media was recorded at 3 weeks. Twenty-one independent  
15       transformants tested were resistant to chlorsulfuron. All but one retained sensitivity to kanamycin.

      The results shown in Table 4 indicate that transformation of tobacco had been achieved with the Agrobacterium harboring the loxP/NptII/Hra Ti plasmid  
20       based on production of chlorsulfuron resistant callus. Since the cells generally remain kanamycin sensitive, these data also suggest that no viable transcript containing the NptII sequence is produced. Nine independent tobacco plants transformed with  
25       loxP/NptII/Hra were assayed for the presence of the loxP-polyA DNA region by Southern blots as described in Materials and Methods. The probe used to hybridize to the filter containing BamHI digested DNA prepared from  
30       each plant was a HindIII-BamHI DNA fragment that was isolated from the pK35K plasmid. This fragment contains the coding region for NptII and is shown in the diagram in Figure 2B. In DNA of each of the nine loxP plants assayed the 2.4 kb BamHI fragment containing the polyadenylation nucleotide sequence and a loxP site, as  
35       diagrammed in Figure 2B, was detected as shown in Figure

2C lanes 1 and 6. None of the loxP plants had the 5.7 kb fragment shown in Figure 2B indicating that no excision could be detected in these primary transformants.

5 In Figure 2B, distances between the BamHI sites and between the loxP sites in the original construction are shown above. Below is a map of the expected configuration following recombination between loxP sites, with the loss of a BamHI site and resulting  
10 change in distance between remaining BamHI sites shown. The 2.4 kb and 5.7 kb fragments marked in bold are those detected by the probe shown as a checkered box.

Plants exhibiting chlorsulfuron resistance were transferred to soil and grown to maturity in a growth  
15 chamber as described above. Individual inflorescences were covered with bags to permit self-fertilization without cross-pollination. Mature seeds were harvested and progeny tests were conducted to determine the inheritance of the inserted DNA fragments. Inheritance  
20 was monitored by following the linked chlorsulfuron resistance trait.

Seed was surface sterilized for 30 minutes as described above, dried, and planted on MX<sup>-</sup> medium in the presence or absence of chlorsulfuron or kanamycin.  
25 Kanamycin was used to assay for the stability of the inactivated NptII gene. Sensitive seeds germinated, but did not develop further. A segregation ratio of 3 chlorsulfuron resistant progeny to 1 sensitive indicated the presence of a single site of integration of the Hra  
30 gene in the genome of the transformant, which was then stably inherited by its progeny. This was seen in 6 out of 20 independent transformants tested. Higher ratios of resistant to sensitive progeny, exhibited by 14 out of 20 of the transformants, indicated insertions at  
35 multiple positions in the genome. For example, a 15/1

ratio indicates the presence of insertions at two unlinked loci and a 255/1 ratio indicates insertions at four unlinked loci in the transformants.

5

TABLE 4

Callus Growth on Chlorsulfuron and Kanamycin

ID		WEIGHT IN GRAMS		
		No	10 mg/l	100 mg/l
		Selection	Chlorsulfuron	Kanamycin
10	Q1	6.99	4.93	0.31
	L1 = Q3	8.19	3.54	0.48
	L2 = Q8	4.18	2.41	0.28
	Q9	2.58	5.78	0.32
	L3 = Q10	6.09	6.04	0.42
15	Q11	9.87	10.19	0.40
	Q12	3.60	4.13	0.31
	Q14	7.57	3.15	0.36
	L4 = Q15	4.69	13.98	0.27
	Q16	3.98	4.92	0.36
20	L5 = Q17	4.92	8.12	0.38
	Q23	5.75	4.94	0.33
	Q25	8.32	8.75	0.31
	Q26	5.30	11.60	1.10
	Q27	5.33	7.38	0.31
25	L6 = Q29	13.28	4.86	0.41
	Q30	6.48	6.52	0.35
	Q31	7.66	12.35	0.31
	Q32	14.41	7.62	0.37
	WT	15.48	0.50	0.90

30

E. Re-transformation of loxP plants with  
the cre coding region

The binary vector plasmid Cre/Hpt-B, described in Example 4, was introduced into transgenic tobacco  
 35 plants, having two loxP sites already integrated in the

genome, by A. tumefaciens infection of tobacco leaf disks from loxP primary transformants. Re-transformed plants were analyzed to demonstrate site-specific recombination at the loxP sites.

5       The procedures described in Materials and Methods were followed, except healthy leaves were harvested from transgenic loxP tobacco plants growing in Magenta GA7 vessels (Magenta Corp., Chicago, IL, USA). These plants were produced as in Section D. Leaf disks, 8 mm in  
10 diameter, were prepared from these axenic leaves using a sterile paper punch. To select for the growth of transformed tobacco cells and to rid the leaf disks of Agrobacterium, a group of leaf disks were transferred to fresh .1N1B medium containing 10-20 mg/L hygromycin and  
15 500 mg/l cefotaxime. To select for site-specific recombination at the loxP sites in transformed tobacco cells and to rid the leaf disks of Agrobacterium, another group of leaf disks were transferred to fresh .1N1B medium containing 100 mg/l kanamycin and 500 mg/l  
20 cefotaxime. Cefotaxime was kept as a frozen 200 mg/ml stock solution and added aseptically (filter sterilized through a 0.45  $\mu$ m filter) to the media after autoclaving. Fresh hygromycin stock (20 mg/ml) and kanamycin stock (50 mg/l) was made for each use and was  
25 filter sterilized into the autoclaved media. Leaf disks were incubated under the growth condition described above for 3 weeks and then transferred to fresh media of the same composition.

Approximately 1 month later, shoots developing on  
30 hygromycin-containing medium were excised with a sterile scalpel and planted in MX<sup>-</sup> medium containing 200 mg/l cefotaxime and 10 mg/l hygromycin in Magenta GA7 vessels. Shoots developing on kanamycin-containing medium were excised with a sterile scalpel and planted



in MX<sup>-</sup> medium containing 200 mg/l cefotaxime and 100 mg/l kanamycin in Magenta GA7 vessels.

Leaves were removed from the rooted excised shoots to determine levels of resistance to hygromycin and kanamycin in a callus induction assay on selective media. To induce callus formation, leaves were excised and leaf disks, 8 mm in diameter, were made using a sterile paper punch and planted on callus induction medium containing 30 mg/l hygromycin and on 50 mg/l kanamycin. Callus growth was recorded at 3 weeks.

The results shown in Figure 2A indicate that Cre-mediated site-specific recombination of the loxP sites in the tobacco genome had been achieved following re-transformation of the loxP plants with the Agrobacterium harboring the Cre/Hpt-B plasmid, based on production of kanamycin resistant callus. Recombination resulting in excision of the polyA site located between the NOS/P and NptII coding region, allowed production of a viable NptII transcript conferring kanamycin resistance to the cells. Only leaf disks from loxP plants re-transformed with the Cre/Hpt-B vector formed callus on medium containing kanamycin. Plants from the transformation of wild type tobacco, with either -/Hpt-B or with Cre/Hpt-B, were all kanamycin sensitive. Three loxP re-transformants from the -/Hpt-B inoculation were also kanamycin sensitive.

In Figure 2A, each bar represents the total weight of five leaf disks grown on callus induction medium (Table 2) containing 50 mg/l kanamycin. Weight includes that of the original leaf disks, which account for weights up to 0.4 grams. Disks were taken from hygromycin selected plants resulting from re-transformation of loxP plants: L1 and L2 or untransformed tobacco plants: WT, with either the cre gene: Cre/Hpt-B vector (i.e., L1\*C/H or L2\*C/H), or

without the cre gene: -/Hpt-B vector (i.e., L1\*-/H or L2\*-/H). All plants exhibited growth on hygromycin.

Plants resulting from re-transformation of loxP plant tissue with Cre/Hpt-B were analyzed by Southern blots to detect recombination. The same NptII fragment probe described in Section D was hybridized to filters containing BamHI digested DNA isolated from re-transformants. Of the five loxP\*Cre re-transformant plant DNAs analyzed, one retained the 2.4 kb fragment detected in loxP primary transformants and in the other four a new 5.7 kb fragment was detected, as shown in Figure 2C lanes 4, 8, 9 and 10. In this figure, lanes contain approximately 10 µg of BamHI digested DNA from the same plants described in (A), in the same order, except that lanes 1 and 6 contain additional samples from the original L1 and L2 plants, respectively. Positions of the 2.4 kb and 5.7 kb bands detected with the NptII probe, as described in (b), are marked. The absence of the 2.4 kb fragment and the presence of the 5.7 kb fragment in these re-transformants indicated that recombination had occurred between the two loxP sites as diagrammed in Figure 2B: excision results in the loss of a BamHI site located between the two loxP sites so that the distance between adjacent BamHI sites is increased. The one loxP\*Cre re-transformant that retained the 2.4 kb fragment, as shown in Figure 2C lane 2, also retained sensitivity to kanamycin, as shown in Figure 2A, demonstrating consistency between the Southern blot analysis and the phenotypic response. DNAs from two control loxP plants re-transformed with -/Hpt contained the 2.4 kb unrecombined fragment as shown in Figure 2C lanes 3 and 7, indicating that recombination is dependent on the presence of Cre. The wild type control plant showed no hybridization to the probe as shown in Figure 2C lane 5.

F. Genetic crosses of heterozygous  
loxP and Cre plants

A method utilized to produce excisional  
5 recombination was to genetically unite the Cre  
recombinase with the loxP-polyA-loxP DNA sequence, which  
is integrated in the plant genome, by sexual  
hybridization of loxP and Cre plants. In this example,  
primary transformants obtained by tobacco leaf disk  
10 transformation, as described in Sections C and D, were  
utilized.

Primary transformants were transferred to soil and  
grown in a growth chamber maintained for a 14 hr, 24°C  
day, 10 hr, 20°C night cycle, with approximately 80%  
15 relative humidity, under mixed cool white fluorescent  
and incandescent lights. Plants were grown to maturity  
and hand pollinations were performed using a slight  
modification of the procedure by Wernsman, E. A. and  
D. F. Matzinger [Hybridization of Crop Plants W. R. Fehr  
20 and H. H. Hadley, eds, pp 657-668 (1980)]. Briefly,  
flowers from Cre plants were selected on the day before  
anthesis; the corolla was split longitudinally, the  
anthers were removed, and the stigma was pollinated with  
pollen from flowers from loxP plants that were allowed  
25 to anthese either on the plant or overnight in a beaker  
of water. To prevent contaminating pollen from reaching  
the stigma, a 4 cm length of a cocktail stirrer, one end  
plugged with modelling clay, was slipped over the stigma  
and style and held in place by the corolla. Each flower  
30 was tagged. Capsules were allowed to grow to maturity  
and then harvested.

The four loxP plants used in the crosses between  
heterozygous parents had segregation ratios that suggest  
the presence of three or more independent loci. All of  
35 the Cre plants used had segregation ratios indicating

insertion of the cre gene at only one genetic locus. Since both the loxP and Cre parents were heterozygous, the seed produced from hand pollinations of loxP pollen onto the stigmas of emasculated flowers from Cre plants could carry none, both, or either one of the foreign DNA insertions. Therefore, seeds resulting from cross pollinations were screened for the presence of the two marker genes and then assayed for kanamycin resistance: a manifestation of site-specific recombination. To identify only those progeny from cross pollinations that carried both markers 100 to 150 seed from each cross were first screened on chlorsulfuron in a germination assay. Then, shoot cuttings of seedlings resistant to chlorsulfuron were tested for root formation in medium containing hygromycin. Seeds from self-crossed Cre and loxP plants were used as controls at each step. Cre x Cre seed produced only bleached seedlings on chlorsulfuron, indicating herbicide sensitivity. None of the loxP x loxP seedlings rooted on hygromycin.

Selected seedlings that were resistant to both compounds, along with controls, were tested for kanamycin resistance using a callus growth assay. A total of 83 out of 90 seedlings (92%) from 8 crosses involving six different Cre and four different loxP parents, were found to be resistant to kanamycin. Table 5 shows the number of seedlings that were kanamycin resistant for each cross. In six of these crosses all of the plants tested were kanamycin resistant (53/53). Progeny from two crosses yielded about 80% kanamycin resistant progeny (20/24 and 10/13). All of the 63 seedlings tested from self-crosses of the loxP and Cre plants were sensitive to kanamycin.

**Table 5**  
**Number of Kanamycin Resistant Progeny**  
**from Crosses Between Heterozygous Cre and loxP Parents**

5	FEMALE	MALE			
		L3 <sup>a</sup>	L4	L5	L6
	self	0/5	0/5	0/5	
	C1 <sup>b</sup>	0/9 <sup>c</sup>	13/13		
	C2	0/9	5/5		
10	C3	0/7	2/2	4/4	16/16
	C4	0/9		20/24	
	C5	0/5	13/13		
	C6			10/13	
15					

<sup>a</sup> L3 - L6 are independent heterozygous loxP transformants used as the pollen parent.

<sup>b</sup> C1 - C6 are different heterozygous Cre transformants used as the female parent.

<sup>c</sup> Number of progeny exhibiting kanamycin resistance/number of progeny tested. Resistance was defined as  $\geq 0.5$  grams callus growth in at least 2 of the 3 leaves tested. Disks were taken from leaves from the distal, middle and proximal portions of the plant to assay for resistance throughout the plant.

Kanamycin resistant plants resulting from genetic crosses of heterozygous loxP plants and Cre plants were assayed by Southern blots to detect recombination. The same NptII fragment probe described in Section D was hybridized to filters containing BamHI digested DNA isolated from progeny of loxP and Cre plants. Of the six progeny DNAs analyzed that were derived from four different crosses involving two loxP plants and four Cre

plants, all contained the 5.7 kb fragment and not the 2.4 kb fragment. DNA of progeny that were controls resulting from selfing of loxP plants retain the 2.4 kb fragment demonstrating that Cre is required for

5 recombination. Another probe was used to verify that the DNA located between the two loxP sites was excised in the progeny of crosses. The BamHI-XbaI fragment containing the polyadenylation nucleotide sequence region of a tobacco Rubisco small subunit gene, that is

10 described in Example 5 and diagrammed in Figure 1C, was labeled and hybridized to the same filters, after washing to remove the first probe. The DNA of progeny of selfed loxP plants contained the 2.4 kb fragment indicating the presence of the polyA sequence while the

15 DNAs of progeny of Cre and loxP plant crosses showed no hybridization to this probe confirming that excision had occurred.

Therefore, both the phenotypic and molecular evidence indicate that Cre mediated site-specific

20 recombination has occurred in these hybrid tobacco seedlings, and the phenotypic data suggests that recombination has occurred in 80% to 100% of the progeny.

25 G. Site-specific recombination in plant cells following genetic crosses of homozygous loxP and Cre plants

A method utilized to produce excisional recombination was to genetically unite the Cre

30 recombinase with the loxP-polyA-loxP DNA fragment, which is integrated in the plant genome, by sexual hybridization of homozygous loxP and Cre plants. Cross pollination using homozygous parents insures the presence of both loxP and cre DNA insertions in all

35 progeny. Plants homozygous for the marker gene linked

to loxP-polyA-loxP and plants homozygous for the marker gene linked to cre were identified and utilized to produce site-specific recombination between two loxP sites.

5        Nine transgenic tobacco plants which had the Hpt marker gene and cre integrated at a single locus, as measured by a 3:1 segregation of the hygromycin resistance in a seed germination assay, were chosen for further crosses. R1 seeds were planted on MX<sup>-</sup> medium  
10        containing 20-50 mg/l hygromycin to select for plants containing the transferred hygromycin resistance gene. Seedlings which were able to develop on the hygromycin-containing medium were transferred to soil and allowed to grow to maturity in the growth chamber maintained for  
15        a 14 hr, 24°C day, 10 hr, 20°C night cycle, with approximately 80% relative humidity, under mixed cool white fluorescent and incandescent lights. Bags were placed on individual inflorescences to permit self-fertilization. Seeds (R2) of several plants (R1)  
20        derived from individual transformants (R0) were collected and subjected to segregation analysis by plating on MX<sup>-</sup> medium containing 50 mg/l hygromycin. R1 plants which were heterozygous would be expected to produce hygromycin resistant progeny with a ratio of  
25        3:1. On the other hand, R1 plants which were homozygous would yield 100% hygromycin resistant progeny after self-fertilization. Using this procedure, homozygous seed stocks of each of the chosen transformants were identified.

30        Five transgenic tobacco plants which had the resistant ALS marker gene and loxP-polyA-loxP integrated at a single locus, as measured by a 3:1 segregation of chlorsulfuron resistance in a seed germination assay, were chosen for further crosses. R1 seed were planted  
35        on MX<sup>-</sup> medium containing 100-300 mg/l chlorsulfuron to

select for plants containing the transferred chlorsulfuron resistance genes. Seedlings which were able to develop on the chlorsulfuron-containing medium were transferred to soil and allowed to grow to maturity in the growth chamber under conditions described above. As above, bags were placed on individual inflorescences to permit self-fertilization. Seeds (R2) of several plants (R1) derived from individual transformants (R0) were collected and subjected to segregation analysis by plating on MX<sup>-</sup> medium containing 300 mg/l chlorsulfuron. R1 plants which were heterozygous would be expected to produce chlorsulfuron resistant progeny with a ratio of 3:1. On the other hand, R1 plants which were homozygous would yield 100% chlorsulfuron resistant progeny after self-fertilization. Using this procedure, homozygous seed stocks of each of the chosen transformants were identified.

Pollen from one homozygous Cre plant, as well as from one WT plant, was used to pollinate three homozygous loxP plants, each derived from an independent primary transformants. To ensure that the hybrid seeds carried both markers, seeds from loxP x Cre crosses were germinated on chlorsulfuron and hygromycin, separately and jointly. All of the germinated seedlings bore true leaves and roots, indicating 100% resistance to both selections. As expected, the progeny of a loxP x WT cross were all chlorsulfuron resistant but hygromycin sensitive.

Hybrid loxP x Cre seeds from the same stocks tested above were germinated and grown on MX<sup>-</sup> medium (Table 2). Tissue from the first through third leaves was tested for growth on kanamycin in a callus induction assay at 25 days after imbibition, and tissue from the fourth through sixth leaves was tested at 40 days after imbibition. The total weight of the leaf disks after



three weeks of callus growth is shown in Figure 3. All seven loxP x WT control progeny tested were sensitive to kanamycin. Most of the 28 progeny from the homozygous loxP x Cre crosses were kanamycin resistant. In two  
5 crosses 100% of the progeny were kanamycin resistant in at least 5 of 6 leaves tested (8/8 and 4/4, respectively). Fourteen out of 16 progeny in another cross were resistant to kanamycin in at least four out of six leaves tested. The two plants which appeared to  
10 be kanamycin sensitive produced more callus than did the controls, but did not show the extent of callus growth associated with resistance (Figure 3).

In Figure 3, C7, L7 and L8 are single locus homozygous plants that were derived from C3, L1 and L2  
15 primary transformants, respectively. L9 is a homozygous loxP plant derived from a primary transformant not used in previous experiments. Each bar represents the weight of 4 leaf disks from an individual offspring, one from each of the first two leaves and two disks from the  
20 third leaf, after incubation on callus induction medium (Table 2) for three weeks. About 0.2 grams is contributed by the original leaf disks. The asterisks mark two progeny that exhibit kanamycin sensitivity.

To determine the number of hybrid progeny in which  
25 site-specific recombination had occurred, the first, second, and third leaves from 94 seedlings from one cross were tested for kanamycin resistance. Seventy one out of the 94 seedlings were kanamycin resistant in all three leaves tested, indicating that recombination had  
30 occurred early in development in 75% of the progeny. However, recombination seems also to have occurred later in development, in that 90% of the seedlings (85/94) exhibited resistance in the third leaf. To assess the incidence of spontaneous kanamycin resistance, 14 self  
35 crossed loxP and 13 self crossed Cre seedlings and 16

loxP x WT progeny were tested in a callus induction assay; none were resistant. This reconfirms that the loxP construction is stable through meiosis, even when the plant carrying it is sexually hybridized.

5

H. Deletion of a sulfonylurea-resistant acetolactate synthase (ALS) marker from transgenic tobacco using the loxP-cre system

EXAMPLE 7

10 A sulfonylurea (SU) resistance marker gene was eliminated from transgenic tobacco plants, leaving the chimeric 35S/P-GUS (GUS= $\beta$ -glucuronidase) gene in the genome. A plasmid was constructed containing a SU-resistant ALS gene located between directly oriented  
15 loxP sites. The vector pTZ19R (Pharmacia, Inc., Piscataway, New Jersey) was digested with HindIII and a synthetic oligonucleotide linker with a nonfunctional HindIII end and XhoI, SalI, HindIII and Asp718 sites was added. This plasmid was digested with HindIII and  
20 ligated with the HindIII fragment from pKNKloxA, described in Example 6 that has directly oriented loxP sites on either end, creating pTZlox2. The resulting plasmid was digested with XbaI, this site being located between the loxP sites, and an XbaI fragment containing  
25 a chimeric SU-resistant ALS gene was added. This chimeric SU-resistant ALS gene is present as an XbaI fragment in a pTZ vector called pMHP35. It contains the CaMV 35S promoter/Cab22L BglIII-NcoI fragment that is described by Harpster et al. [Mol. Gen. Genet. 212: 182-  
30 190 (1988)] and the Arabidopsis ALS coding and 3' regions, described by Mazur et al. [Plant Physiol. 85: 1110-1117 (1987)], that was mutated so that it encodes a SU-resistant form of ALS. The mutations, introduced by site directed mutagenesis, are those present in the  
35 tobacco SU-resistant Hra gene described by Lee et al.

[EMBO J. 5: 1241-1248 (1988)]. The resulting plasmid in which the SU-resistant ALS gene is between loxP sites was named pTZlox2FA. Next the entire lox-ALS-lox fragment was isolated following SalI and Asp718

5 digestion, and cloned into the binary vector pZS94 that had been digested with the same enzymes creating pZ4loxA.

pZS94 contains the origin of replication and ampicillin resistance gene from pBR322 for maintenance  
10 and selection in E. coli. It contains the replication and stability regions of the Pseudomonas aeruginosa plasmid pVS1, described by Itoh et al. [Plasmid 11: 206-220 (1984)], which are required for replication and maintenance of the plasmid in Agrobacterium. Also  
15 contained are a T-DNA left border fragment of the octopine Ti plasmid pTiA6 and a right border fragment derived from TiAch5 described by van den Elzen et al. [Plant Molec. Biol. 5: 149-154 (1985)]. Between these borders are a LacZ gene and the unique restriction sites  
20 HindIII, SalI, BamHI, SmaI, Asp718, and EcoRI derived from pUC18. pZ4loxA was digested with SalI and a SalI fragment containing a chimeric 35S/P-GUS gene was added. This chimeric GUS gene contains the 35S promoter/Cab22L fragment described above, the GUS coding region  
25 available from Clontech, and the Nos 3' region described in Example 1. The resulting plasmid was named pZ4loxAG and is shown in Figure 4.

In Figure 4, the pZS94 binary vector contains a chimeric 35S/P-ALS gene that is bounded by directly  
30 oriented loxP sites, and a chimeric 35S/P-GUS-Nos 3' gene. The loxP sites are indicated by arrowheads.

pZ4loxAG was transferred into A. tumefaciens LBA4404 by direct DNA uptake following the procedure described in Plant Molecular Biology Manual, [SB Gelvin et al.,  
35 eds. Kluwer Academic Press PMAN-A3/7, (1988)]. The

presence of the binary vector in Agrobacterium colonies selected on minA medium with sucrose (See Table 1) containing 100 µg/ml carbenicillin was verified by restriction digests of miniprep DNA. The resulting  
5 Agrobacterium strain was used to obtain tobacco transformants as described in Materials and Methods using resistance to 25 ppb chlorsulfuron for selection of transformants, as detailed below.

Leaf disks were inoculated by submerging them for  
10 several minutes in 20 ml of a 1:20 dilution of the overnight culture of Agrobacterium. The culture was started by inoculating 5 ml of YEP medium (Table 8) containing 100 mg/l carbenicillin with a single bacterial colony. The culture was grown for  
15 approximately 17-20 hours in a glass culture tube in a New Brunswick platform shaker maintained at 28°C.

After inoculation, the leaf disks were placed in petri dishes containing .1N1B agar medium (Table 2) and sealed with parafilm. The petri dishes were incubated  
20 under mixed fluorescent and "Gro and Sho" plant lights (General Electric) for 2-3 days in a culture room maintained at approximately 25°C.

To rid the leaf disks of Agrobacterium and to select for the growth of transformed tobacco cells, the leaf  
25 disks were transferred to fresh .1N1B medium containing 500 mg/l cefotaxime and 25 ppb chlorsulfuron. Cefotaxime is kept as a frozen 200 mg/ml stock solution and added aseptically (filter sterilized through a 0.45 µm filter) to the media after autoclaving. A  
30 frozen chlorsulfuron stock was prepared by first making a 0.2 mg/ml solution in 0.01 N NH<sub>4</sub>OH, which was then diluted 1:10 with deionized water, and filter sterilized into the autoclaved media. Leaf disks were incubated under the growth condition described above for 18 days

and then transferred to fresh media of the same composition.

Fifteen days later, shoots developing on medium containing 25 ppb chlorsulfuron were excised with a sterile scalpel and planted in MX<sup>-</sup> medium containing 500 mg/l cefotaxime and 25 ppb chlorsulfuron. Root formation was recorded within 2 weeks.

Integration of an intact GUS/loxP/Hra DNA sequence in the plant genome was verified by a callus induction assay for chlorsulfuron resistance and a GUS enzyme activity assay. A leaf piece, approximately 0.2 cm<sup>2</sup> in size, was removed from each of several selected rooted excised shoots to test for GUS activity. Presence of GUS activity was determined by grinding each leaf piece in a solution (Table 6) containing 1 mg/ml X-Gluc (5-bromo-4-chloro-3-indolyl- $\beta$ -gulcuronide) and incubating the tissues 1-16 hours at 37°C. The formation of a blue precipitate indicated the presence of GUS activity.

TABLE 6  
Solution for X-Gluc

	<u>Stocks</u>	<u>Volume (ml)</u>
25	0.2 M NaPO <sub>4</sub> buffer, pH 7.0 (0.2 M Na <sub>2</sub> HPO <sub>4</sub> : 62 ml 0.2 M NaH <sub>2</sub> PO <sub>4</sub> : 38 ml)	25.0
	Deionized water	24.0
	0.1 M K <sub>3</sub> [Fe(CN) <sub>6</sub> ]	0.25
30	0.1 M K <sub>4</sub> [Fe(CN) <sub>6</sub> ]·3H <sub>2</sub> O	0.25
	1.0 M Na <sub>2</sub> EDTA	0.50

Leaves from five selected plants exhibiting GUS activity were removed from the rooted excised shoots to determine levels of resistance to chlorsulfuron in a

callus induction assay on selective media. To induce callus formation, leaves were excised and 10 leaf disks, 8 mm in diameter, were made using a sterile paper punch and were plated on callus induction medium containing 5 25 ppb chlorsulfuron and 250 µg/l cefotaxime. Wild-type was used as a control. Leaf disks and the associated callus was weighed; all five transformants exhibited resistance to chlorsulfuron as shown in Table 7.

To test for excision of the ALS marker gene, the 10 five chlorsulfuron-resistant independent transformants that exhibited GUS activity were re-transformed to introduce the cre gene as follows. Healthy leaves were harvested from these five transgenic tobacco plants and a wild-type plant growing in Magenta GA7 vessels 15 (Magenta Corp., Chicago, IL, USA). Leaf disks were prepared from these axenic leaves using a sterile paper punch, inoculated with Agrobacterium harboring either the -/Hpt or the Cre/Hpt-B plasmid (Example 4), incubated for three days on .1N1B, placed on .1N1B 20 medium containing 500 mg/l cefotaxime and 30 mg/l hygromycin, and incubated 2 weeks. Leaf disks were then transferred to fresh medium of the same composition for two weeks, then transferred to MX- medium containing 500 mg/l cefotaxime and 30 mg/l hygromycin for shoot 25 formation. As shoots appeared, they were excised from the leaf disk and placed in MX- medium containing 500 mg/l cefotaxime and 30 mg/l hygromycin. Shoots were taken from different leaf disks to ensure that they result from independent transformation events. Shoots 30 which rooted were assayed for GUS activity using X-Gluc as described above. Sixty-three shoots tested had GUS enzyme activity while three did not exhibit GUS activity. The parent pZ4loxAG plant of these three retransformants appears to be chimeric for GUS 35 expression.

Leaf disks were taken from each plant and tested in a callus induction assay as described above. Results are shown in Table 7. Ninety-one percent of all plants resulting from re-transformation with -/Hpt remained chlorsulfuron resistant, as expected, since the Cre recombinase was not introduced. Ninety-five percent of the plants resulting from re-transformation of pZ4LoxAG plants with Cre/Hpt-B exhibited sensitivity to chlorsulfuron; only two out of 38 plants remained resistant to chlorsulfuron. These data indicate that the sulfonylurea-resistant ALS gene is no longer functioning in a majority of those plants which received the cre coding region, suggesting that it has been excised by cre-mediated recombination at loxP sites.

To verify that excision has occurred, DNA prepared from re-transformants was analyzed on Southern blots. The DNA was digested with EcoRI before gel electrophoresis and transferred to filters. A DNA fragment consisting of the 35S promoter was prepared, radioactively labeled, and hybridized to the blotted plant DNA. A 6 kb band was detected in re-transformants that did not receive cre. This band represents a DNA fragment extending from an EcoRI site located 3' to the GUS coding region to an EcoRI site located within the ALS coding region (see Figure 4). It includes the GUS coding region, the 35S promoter that regulates expression of GUS, the 35S promoter that regulates expression of ALS and a portion of the ALS coding region. A 3.2 kb band was detected in DNA from plants that received cre. This band represents a DNA fragment extending from the EcoRI site located 3' to the GUS coding region to an EcoRI site located just outside of the distal loxP site. It includes the GUS coding region and the 35S promoter regulating GUS expression. The shift of the detected fragment size from 6 kb to 3.2 kb

verifies the excision of the DNA segment located between the loxP sites, including the ALS coding region and the 35S promoter that regulates its expression.

5

TABLE 7

Number of Chlorsulfuron Resistant and Sensitive<sup>1</sup> Plants Recovered after Re-transformation with -/Hpt and Cre/Hpt-B as Determined by a Callus Induction Assay

10	Original		Re-transformed			
	Transformant		Hpt/-		Cre/Hpt-B	
			#	#	#	#
	ID	Resistant?	Resistant	Sensitive	Resistant	Sensitive
	D1	yes	7	0	2	11
15	D2	yes	4	0	0	5
	D3	yes	9	0	0	8
	D4	yes	6	0	0	8
	D5	yes	4 <sup>2</sup>	2	0	4
	WT	no	0	1		

20

<sup>1</sup> Resistant is defined as average weight per leaf disk  $\geq$  1.0 gram and sensitive is defined as average weight per leaf disk  $\leq$  0.2 grams.

<sup>2</sup> One of these transformants was resistant in one  
25 experiment and sensitive in another experiment.

## EXAMPLE 8

Deletion of a SU-resistant ALS marker gene from transgenic arabidopsis using the loxP-cre system

30 A sulfonylurea (SU) resistance marker gene was eliminated from transgenic Arabidopsis plants, leaving the chimeric 35S/P-GUS gene in the genome. Plasmid construction and transformation of Agrobacterium are described in Example 7.



Standard aseptic techniques for the manipulation of sterile media and axenic plant/bacterial cultures are followed, including the use of a laminar flow hood for all transfers. Compositions of the culture media are listed in Table 8. Unless otherwise indicated, 25x100 mm petri plates, sealed with filter tape (Carolina Biological Supply Co., Burlington, NC), were used for plant tissue cultures. Incubation of plant tissue cultures was at 23°C under constant illumination with mixed fluorescent and "Gro and Sho" plant lights (General Electric) unless otherwise noted.

The source of explants was in vitro grown roots of Arabidopsis thaliana (L.) Heynh, geographic race Wassilewshija. Seeds were sterilized for 10 min in a solution of 50% commercial bleach with 0.1% SDS, rinsed three to five times with sterile water, dried thoroughly on sterile filter paper, and then 2-3 seeds were sown in 50 ml liquid Gamborg's B5 medium (Gibco #560-1153) in 250 ml Belco flasks. The flasks were capped, placed on a rotary shaker at 70-80 rpm, and incubated for 3-4 weeks.

Prior to inoculation with Agrobacterium, root tissues were cultured on callus induction medium (MSKig, infra). Roots were harvested by removing the root mass from the Belco flask, placing it in a petri dish, and using forceps, pulling small bundles of roots from the root mass and placing them on MSKig medium. Petri dishes were sealed with filter tape and incubated for four days.

Cultures of Agrobacterium cells containing the binary plasmid pZ4loxAG, as previously described, were grown in 5 ml of YEP medium containing 100 mg/l carbenicillin. Cultures of Agrobacterium cells containing the binary plasmid Cre/Hpt-B (Example 3) were grown in 5 ml of YEP medium containing 5 mg/l

tetracycline. The cultures were grown for approximately 17-20 hours in glass culture tubes in a New Brunswick platform shaker (225 rpm) maintained at 28°C. Pre-cultured roots were cut into 0.5 cm segments and placed

5 in a 100  $\mu$ m filter, made from a Tri-Pour beaker (VWR Scientific, San Francisco, CA USA) and wire mesh, which is set in a petri dish. Root segments were inoculated for several minutes in 30-50 ml of a 1:20 dilution of the overnight Agrobacterium culture with periodic gentle

10 mixing. Inoculated roots were transferred to sterile filter paper to draw off most of the liquid. Small bundles of roots, consisting of several root segments, were placed on MSKig medium containing 100  $\mu$ m Acetosyringone (3',5'-Dimethoxy-4'-hydroxyaceto-phenone,

15 Aldrich Chemical Co., Milwaukee, WI, USA). Petri plates were sealed with parafilm or filter tape and incubated for two to three days.

After infection, root segments were rinsed and transferred to shoot induction medium with antibiotics

20 as detailed below. Root bundles were placed in a 100- $\mu$ m filter unit (described above) and rinsed with 30-50 ml liquid MSKig medium. The filter was vigorously shaken in the solution to help remove the Agrobacterium, transferred to a clean petri dish, and rinsed again.

25 Roots were blotted on sterile filter paper and bundles of roots were placed on MSg (infra) medium containing 500 mg/l vancomycin and either 25 ppb chlorsulfuron (pZ4loxAG) or 15 mg/l hygromycin (Cre/Hpt and -/Hpt). Plates were sealed with filter tape and incubated for 12

30 to 14 days.

Green nodules and small shoot primordia were visible at about 2-3 weeks. The explants were either left intact or were broken into numerous pieces and placed on GM medium containing 200-300 mg/l vancomycin and either

35 25 ppb chlorsulfuron (pZ4loxAG) or 10 mg/l hygromycin

(Cre/Hpt and -/Hpt) for further shoot development.

Plates were either sealed with two pieces of tape or with filter tape. As they developed, individual shoots were isolated from the callus and were placed on MSRG

- 5 medium containing 100 mg/l vancomycin and either 25 ppb chlorsulfuron (pZ4loxAG) or 10 mg/l hygromycin (Cre/Hpt and -/Hpt). Dishes were sealed as described above and incubated for seven to 10 days. Shoots were then transferred to GM medium containing 100-200 mg/l
- 10 vancomycin in 25x100 petri dishes or Magenta G7 vessels. Many primary transformants (T1) which were transferred to individual containers set seed (T2).

- T2 seed was harvested from selected putative transformants and sown on GM medium containing either 25
- 15 ppb chlorsulfuron (pZ4loxAG) or 10-30 mg/l hygromycin (Cre/Hpt and -/Hpt). Plates were cold treated for 2 or more days at 4°C, and then incubated for 10 to 20 days at 23°C under constant illumination as described above. Seedlings were scored as resistant (green, true leaves
- 20 develop) and sensitive (no true leaves develop).

Selected chlorsulfuron or hygromycin resistant T2 seedlings were transplanted to soil and were grown to maturity at 23°C daytime (14 hours), 18°C nighttime (10 hours), at 65-80% relative humidity.

- 25 Genetic crosses of Cre/Hpt-B plants (hygromycin resistant) and pZ4loxAG plants (chlorsulfuron resistant) were performed and the resulting seed allowed to mature. Seed was collected, sterilized, plated on GM (Table 8) containing 30 mg/l hygromycin, and tissue from the
- 30 seedlings tested for GUS activity with X-gluc as previously described. Those seedlings exhibiting both hygromycin resistance and GUS activity (indicating that they received both the cre gene and the loxP construction), were allowed to grow until stem tissue
- 35 could be obtained for a callus induction assay. Stem

tissues were obtained from selected transformants and placed on MSKig medium with and without 25 mg/l chlorsulfuron. Callus growth was recorded within three weeks. pZ4LoxAG plants were allowed to self-pollinate, seed was collected, sterilized, plated on GM containing 25 ppb chlorsulfuron, and used as controls (see Table 9 for results). Sensitivity to chlorsulfuron in the majority of GUS positive seedlings resulting from crosses between Cre/Hpt-B and pZ4LoxAG plants indicates that in these seedlings, the SU-resistant ALS marker gene was no longer functioning, suggesting it has been excised by Cre-mediated recombination between loxP sites.

Chlorsulfuron sensitive plants were placed in soil and are allowed to mature. T3 seed is collected, sterilized, and germinated on GM medium with or without 30 mg/l hygromycin or 25-100 ppb chlorsulfuron. Plates are sealed with filter tape, cold treated for 2 or more days at 4°C, and then incubated for 10 to 20 days at 23°C under constant illumination as described above. Seedlings are scored as resistant and sensitive and the results recorded. Representative seedlings are screened for GUS activity. Some seedlings exhibit GUS activity, but are not resistant to chlorsulfuron indicating that the SU-resistant ALS marker gene is no longer functioning, suggesting that it has been excised by Cre-mediated recombination between loxP sites.

To verify that excision has occurred, DNA prepared from these plants is analyzed on Southern blots. The DNA is digested with EcoRI before gel electrophoresis and transferred to filters. A DNA fragment consisting of the 35S promoter is prepared, radioactively labeled, and hybridized to the blotted plant DNA. A 6 kb band is detected in re-transformants that did not receive cre. This band represents a DNA fragment extending from an

EcoRI site located 3' to the GUS coding region to an EcoRI site located within the ALS coding region (see Figure 4). It includes the GUS coding region, the 35S promoter that regulates expression of GUS, the 35S promoter that regulates expression of ALS, and a portion of the ALS coding region. A 3.2 kb band is detected in DNA from plants that received cre. This band represents a DNA fragment extending from the EcoRI site located 3' to the GUS coding region to an EcoRI site located just outside of the distal loxP site. It includes the GUS coding region and the 35S promoter regulating GUS expression. The shift of the detected fragment size from 6 kb to 3.2 kb verifies the excision of DNA segment located between the loxP sites, including the ALS coding region and the 35S promoter that regulates its expression.

TABLE 8  
Medium Composition

20

YEP MEDIUM

	<u>Per Liter</u>
Bacto Peptone	10.0 g
Bacto Yeast Extract	10.0 g
NaCl	5.0 g
Agar (optional)	15.0 g
pH 7.0	

BASIC MEDIUM

- 25 1 pkg. Murashige and Skoog Minimal Organics Medium  
without Sucrose (Gibco #510 or Sigma # M6899)  
10 ml Vitamin Supplement  
0.05% MES 0.5 g/l  
0.8% agar 8 g/l  
30 pH 5.8

VITAMIN SUPPLEMENT - 100 X Stock

10 mg/l thiamine  
 50 mg/l pyridoxine  
 5 50 mg/l nicotinic acid

GM = Germination Medium

Basic Medium

1% sucrose 10 g/l

10

MSKig = Callus Induction Medium

Basic Medium

2% glucose 20 g/l  
 0.5 mg/l 2,4-D 2.3  $\mu$ M  
 15 0.3 mg/l Kinetin 1.4  $\mu$ M  
 5 mg/l IAA 28.5  $\mu$ M

MSg = Shoot Induction Medium

Basic Medium

20 2% glucose 20 g/l  
 0.15 mg/l Indole-3-Acetic Acid 0.86  $\mu$ M  
 (IAA)  
 5.0 mg/l N<sup>6</sup>-( $\Delta$ <sub>2</sub> Isopentenyl)- 24.6  $\mu$ M  
 Adenine 2iP

25

MSRg = Shoot Induction Medium

Basic medium

2% glucose 20 g/l  
 12 mg/l Indole-3-Butyric Acid 58.8  $\mu$ M  
 30 (IBA)  
 0.1 mg/l Kinetin 0.46  $\mu$ M

TABLE 9

Results of Assays of Seedlings Resulting from Crosses  
between pZ4loxAG and Cre/Hpt-B Arabidopsis Plants

5	# Chlorsulfuron Sensitive Seedlings/ # GUS Positive, Hygromycin Resistant	
	<u>Female</u>	<u>Male</u>
		<u>Seedlings Tested</u>
	E1	B6
	E2	B4
10	E3	B3
	E4	B5
	E5	B3
		42/42
		0/5
		4/4
		7/7
		1/1

15

EXAMPLE 9

Expression of Cre from the chemically  
regulated corn promoter In2-2

Cre expression was placed under control of the In2-2 promoter, that is induced by N-(aminocarbonyl)-2-chlorobenzenesulfonamide, by constructing the chimeric gene: In2-2/P-cre-Nos 3'. The starting material for the construction was the plasmid Cre/Hpt-A, which was described in Examples 1 and 2. Cre/Hpt-A was digested with HindIII and SalI and the DNA fragment containing the cre coding region and the Nos 3' was isolated. This HindIII-SalI fragment was subcloned into the vector Bluescript SK(+) (Stratagene, catalog #212205) that had been digested with HindIII and SalI and dephosphorylated, yielding the plasmid designated pBSCre.

The addition of the In2-2 promoter to pBSCre was accomplished using plasmids HPH 463 dam(-) and 2-2(3.9) which have been described in WO 90/11361. The plasmid 2-2(3.9) is a pUC18 plasmid containing a 3.9 kb SalI fragment derived from the genomic clone containing the

2-2 gene. The 3.9 kb fragment includes 3.6 kb of promoter sequence located 5' of the translation start site and 180 bp of the coding region for the 2-2 protein. The plasmid pHPH 463 dam(-) is a Bluescript S/K(+) plasmid containing a chimeric promoter that has 136 bp of 2-2 promoter joined to the 5' untranslated leader from the maize alcohol dehydrogenase (ADH) 1-1S allele [Dennis et al., Nucl. Acids Res. 12: 3983-4000 (1984)], with an NcoI site incorporated at the translation start codon. HPH 463 dam(-) was digested with ClaI and the resulting 5' overhangs were filled in with Klenow. The resulting DNA was then digested with XbaI and the ClaI blunt-XbaI fragment containing the 3' part of the 2-2 promoter and maize ADH leader was isolated. The plasmid pBSCre was digested with HindIII and the resulting 5' overhangs were rendered blunt with Klenow. This pBSCre DNA was then digested to completion with XbaI (located in the polylinker) and dephosphorylated using calf intestinal phosphatase. The ClaI blunt-Xba I fragment derived from HPH463 dam(-) and the dephosphorylated pBSCre vector were then ligated together to yield the plasmid designated pBSCre101. This construction contains a cre coding region under the transcriptional control of a modified 2-2 promoter that includes a 5' untranslated leader sequence from the maize ADH gene.

Next the 5' distal portion of the In 2-2 promoter was added to pBSCre101. The plasmid 2-2(3.9), described above, was digested to completion with XbaI and AatII and a fragment containing 1.2 kb of the 2-2 promoter was isolated. The plasmid pBSCre101 was digested with XbaI and AatII, dephosphorylated and ligated to the XbaI-AatII fragment from 2-2(3.9). The resulting construction was designated pBSCre102.



Next pBSCre102 was digested with XbaI and XhoI and the fragment containing the entire In2-2/P-cre-Nos 3' chimeric gene was ligated to pJJ2644 that had been digested with XbaI and XhoI and dephosphorylated, resulting in pBS103. pJJ2644 was described in Example 4. pBSCre103, shown in Figure 5, was transformed into A. tumefaciens and the resulting strain used to obtain tobacco transformants as described in Materials and Methods; 30 mg/l hygromycin was used for selection and leaf disks were placed on fresh medium every two to three weeks.

Independent primary transformants were grown in magenta boxes, and young leaves were harvested and frozen in liquid nitrogen. One week after the initial harvest, shoot tips were harvested and placed in 15-ml Falcon tubes filled with 0.5X Hoagland's solution (Table 11) containing 200 mg/l N-(aminocarbonyl)-2-chlorobenzenesulfonamide. The plants were allowed to take up inducer for approximately 24 hours. The induced shoots were then harvested, frozen in liquid nitrogen and stored at -80°C.

Total RNA was isolated from both sets of samples as described by Colbert et al. [Proc. Natl. Acad. Sci. USA 80: 2248-2252 (1983)]. Replicate RNA samples from both uninduced and induced transformed plants were transferred to nitrocellulose filters and probed with nick translated pBSCre. Twenty-six primary transformants had little to no cre mRNA in the uninduced state, and a strongly hybridizing cre mRNA signal after treatment with N-(aminocarbonyl)-2-chlororbenzene-sulfonamide. This result demonstrates that expression of the cre gene was successfully regulated when under control of the In2-2 promoter.

TABLE 100.5X Hoagland's Nutrient Solution

5	1.0 mM ammonium phosphate, monobasic
	4.0 mM potassium nitrate
	4.0 mM calcium nitrate
	2.0 mM magnesium sulfate
	1.0 mM ammonium nitrate
	5.0 ppb Sequestrene
10	9.2 µM manganese chloride
	46.0 µM boric acid
	0.77 µM zinc sulfate
	0.32 µM cupric sulfate
15	0.11 µM molybdic acid

EXAMPLE 10

Chemical regulation of loxP-cre  
mediated recombination

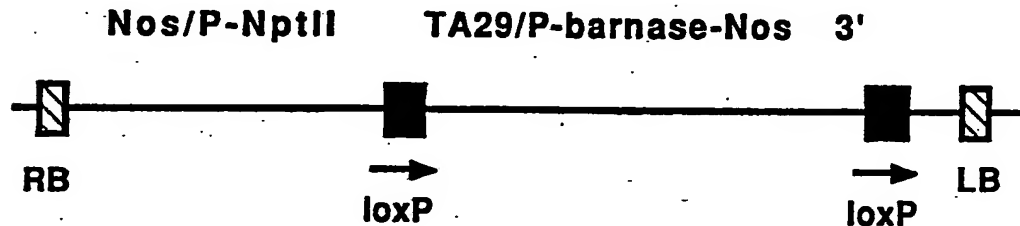
Transformants containing the In2-2/P-cre-Nos 3' gene  
 20 that respond to induction by N-(aminocarbonyl)-2-  
 chlorobenzenesulfonamide, described in Example 9, were  
 crossed, as described in Section F, with the homozygous  
loxP plants that were described in Section G. The  
 resulting seed is germinated on hygromycin to select  
 25 those progeny that receive the cre gene. All progeny  
 receive the loxP construction from the homozygous  
 parent. Leaf disks are taken from selected progeny and  
 induced to callus on kanamycin medium as described  
 previously. Lack of callus formation indicates that the  
 30 lox construction, in which the kanamycin resistance gene  
 is nonfunctional due to the loxP-polyA-loxP fragment  
 insertion, remains intact and that no recombination has  
 occurred. Cre expression is also induced in progeny by  
 treating with N-(aminocarbonyl)-2-chlorobenzene-  
 35 sulfonamide or related inducing chemicals by spraying,

by cutting off the shoot tip or a leaf and allowing uptake through the vascular system, or by placing tissue explants on inducer-containing medium. Following induction, leaf disks are again assayed for callus growth on kanamycin. Growth of this tissue indicates that Cre-mediated recombination at the loxP sites excises the polyA fragment and restores function of the kanamycin resistance gene. Thus recombination occurs only after the inducing chemical is applied to plants containing the In2-2/P-cre-Nos 3' gene and a loxP construction.

#### EXAMPLE 11

##### Restoration of fertility in a molecular genetic approach to hybrid seed production

An anther promoter-cell disruption chimeric gene (A/P-CD) causing male sterility is deleted from the genome of the hybrid plant resulting in fertility restoration. The A/P-CD gene used consists of the TA29 promoter, the barnase coding region, and the NOS 3' end as described in EPA 89-344029. A binary vector is constructed to consist of the A/P-CD gene flanked by directly repeated loxP sites, and a chimeric NptII gene as a kanamycin resistance selection marker, between the left and right T-DNA border sequences. A diagram of the T-DNA of the binary vector is shown below:



This plasmid, called Lox/AD, is transferred into A. tumefaciens LBA4404 and the resulting strain is used

to obtain tobacco transformants as described previously. The resulting transformants containing Lox/AD are grown to maturity and tested for male sterility, which indicates correct expression of the A/P-CD gene. Male sterile Lox/AD plants are identified as those that produce no seed upon selfing and/or produce no pollen. These plants are fertilized with pollen from homozygous 35S/P-Cre plants, obtained in Section C, and seed is harvested as previously described. Seeds are sterilized and planted on MX- medium in the presence of kanamycin. Seedlings which are resistant to kanamycin, and therefore have received the Lox/AD construction, are grown to maturity and tested for male fertility. Restoration of fertility is identified by the ability to develop pollen and produce seed upon selfing. Fertility restoration indicates that the A/P-CD gene is deleted due to interaction of Cre protein with the loxP sites.

#### EXAMPLE 12

##### Cre-lox mediated disruption of F1 seed development

F1 seed development is aborted by activating a seed disruption gene using the loxP-Cre system. One component of a seed disruption gene is a promoter that is only expressed in the seed. This type of promoter can be derived from a gene whose expression is naturally associated with the embryo and/or endosperm. Desirable promoters to use are derived from the embryo-expressed  $\beta$ -subunit of phaseolin ( $\beta$ -Ph), or from the  $\alpha'$  subunit of  $\beta$ -conglycinin of soybean ( $\alpha'$ - $\beta$ -CG), which is highly expressed early in seed development in the endosperm and embryo. These two genes are described by Doyle et al. [J. Biol. Chem., 261:9228-9238 (1986)].

A second component of the seed disruption gene is a coding region which produces a protein that disrupts

normal cell functions. An example is the coding region for barnase derived from Bacillus amyloliquefaciens, which has been cloned and characterized by Hartley [J. Mol. Biol. 202: 913-915 (1988)] and used in EPA 89-344029. The third component is a polyadenylation signal sequence region, which can be derived from the 3' end of most any gene that is functional in plant cells. In this example we use the 3' region from the bean phaseolin gene [Chee et al., Gene 41: 457 (1986)]. The seed disruption gene is made into an inactive form by placing a loxP-poly-A-loxP DNA fragment between the promoter and coding region as described for the NOS/P-NptII gene in Example 5. Inactive and active (control) chimeric genes containing either the  $\beta$ -Ph or the  $\alpha'$ - $\beta$ -CG promoter, the barnase coding region (bar), and the phaseolin 3' region (Ph 3') are constructed as follows.

Plasmids containing the  $\beta$ -Ph promoter and Ph 3' or the  $\alpha'$ - $\beta$ -CG promoter and Ph 3' called pUC18pvPpvS and pUC18gmPpvS, respectively, were obtained from Dr. Jerry Slightom, the Upjohn Company. The promoter and 3' regions contained in these two plasmids were synthesized from the genes described by Doyle et al. [J. Biol. Chem., 261:9228-9238 (1986)] using the polymerase chain reaction (PCR) procedure described by Saiki et al. [Science, 239:487-491 (1987)]. During the PCR procedure an NcoI site was added at the translation start ATG and 5' to the Ph 3' sequence by incorporating the NcoI restriction site recognition sequence into the appropriate synthetic oligonucleotide primers. A HindIII site was similarly added at the 5' end of the promoter fragments. The synthetic promoter and 3' region fragments were joined at the introduced NcoI sites and ligated into the HindIII site of pUC18 (a HindIII site occurs naturally at the end of the Ph 3'

fragment). The resulting plasmids, pUC18pvPpvS and pUC18gmPpvS, were each digested with EcoRI and SalI, the ends filled in using the Klenow enzyme, and religated to delete the polylinker sites located between EcoRI and SalI. The resulting plasmids were named CW104 and CW105, respectively. Each of these plasmids was then digested with NcoI and a synthetic oligonucleotide with restriction sites: NcoI, SmaI, KpnI, XbaI, and incomplete NcoI was added. The resulting plasmids were named CW108 and CW109.

Since the barnase enzyme is lethal to cells, an inhibitor of barnase called barstar is expressed in the same cells. The pMT420 plasmid containing the barstar and barnase genes isolated from Bacillus amyloliquefaciens was obtained from Dr. Robert Hartley, NIH. These genes are described by Hartley [J. Mol. Biol. 202:913-915 (1988)]. The barstar gene was isolated from pMT420 as a PstI-HindIII fragment and ligated into PstI (a unique PstI polylinker site is located after the HindIII site at the 3' end of the Ph 3' fragment) and HindIII digested CW108. A plasmid that retains the HindIII  $\beta$ -Ph promoter and Ph 3' fragment and contains the barstar gene was identified and named 108B. The NcoI-PstI fragment containing Ph 3' and barstar was isolated and ligated to NcoI and PstI digested CW109 creating 109B.

The complete barnase protein includes a pre-sequence involved in secretion, a pro-sequence involved in folding, and the mature protein sequence containing the enzyme activity. We propose that expression of only the mature protein is most effective in disrupting plant cells. Though EPA 89-344029 makes use of barnase as a plant cell disruption protein, it does not disclose details on construction of a barnase gene for expression

in plant cells. No information is given on the portion of the barnase coding region that is expressed.

To prepare a DNA fragment containing the coding region for the mature barnase protein (bar), *Bacillus* 5 *amyloliquefaciens* DNA was used as a template for the PCR procedure described by Saiki et al. [Science 239:487-491 (1987)]. Synthetic primers were made that add an NcoI site, including an in-frame translation start ATG, at the 5' end of the mature protein coding region and an 10 XbaI site following the translation stop codon. The amplified DNA fragment was digested with NcoI and XbaI and ligated to NcoI and XbaI digested 108B creating 108BB, which then contains an active form of the seed disruption gene. Similarly, 109BB is constructed.

15 To prepare inactive forms of the seed disruption gene, first the NcoI site at the 3' end of the promoter fragment was removed from CW108 and CW109, described above, by digesting each plasmid with NcoI and treating with S1 nuclease. S1 treated CW108 was religated, a 20 plasmid missing the NcoI site was identified by restriction mapping and DNA sequencing, and named 108N. S1 treated CW109 was digested with SmaI (the site is adjacent to the NcoI site) and religated. A plasmid missing the NcoI site was identified by restriction 25 mapping and DNA sequencing and named 109N1. 109N1 was digested with Asp718 and XbaI and a synthetic oligonucleotide linker with the sites Asp718, XhoI, NcoI, and XbaI was added. The resulting plasmid named 109N1X was digested with XhoI and NcoI and an XhoI-NcoI 30 loxP-polyA-loxP DNA fragment was added. This fragment was prepared from p69ssN, a derivative of pBS69polyA, which was described in Example 5. To make p69ssN, pBS69polyA was digested at the HindIII site located outside of one loxP site, the ends filled in, and NcoI 35 linkers were added. 109N1X containing the loxP-polyA-

loxP fragment was called 109lox2. This plasmid is digested with NcoI and PstI and the NcoI-PstI fragment prepared from 109BB containing bar, Ph 3', and barstar is added creating 109lox2BB. This plasmid contains an inactive seed disruption gene. 108N is digested with Asp718 and PstI and the Asp718-PstI fragment prepared from 1098lox2BB is added. The resulting plasmid is named 108lox2BB and contains an inactive seed disruption gene.

10 The inactive and active cell disruption genes are each moved into a binary vector with an NptII gene within the borders and a barstar gene added outside of the T-DNA borders. The resulting plasmids are called pZ108lox2BB, pZ109lox2BB, pZ108BB, and pZ109BB. These  
15 plasmids are transferred into a disarmed A. tumefaciens and the resulting strains are used to obtain transformants as previously described.

Expression of the cre coding region is more effective either with the same developmentally  
20 controlled promoter or with the highly active 35S promoter. The cre coding region was placed under control of the same seed promoters used for the disruption genes making chimeric SP-cre-Ph 3' genes: 108Cre and 109Cre. These genes were cloned into binary  
25 vectors between the T-DNA borders along with a chimeric sulfonylurea resistance selection marker gene, creating pZ108Cre and pZ109Cre plasmids. These plasmids are transferred into a disarmed A. tumefaciens and the resulting strains are used to obtain tobacco or  
30 Arabidopsis transformants as previously described. Homozygous single locus plants are derived from primary transformants as previously described.

Homozygous plants transformed with pZ108lox2BB or pZ109lox2BB are crossed with homozygous plants  
35 containing pZ108Cre or pZ109Cre, and with the homozygous



Cre plants described in Section C or Example 9. Seed pods or siliques are checked for the absence of seed indicating that the seed disruption gene is activated by the loxP-cre system.

5

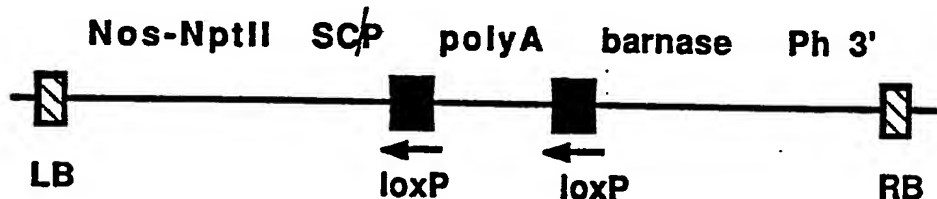
### EXAMPLE 13

#### Cre-lox mediated disruption of F2 seed development

F1 seed development is normal and F2 seed development is aborted by activating a seed coat disruption gene using the loxP-cre system. A seed coat disruption gene is made consisting of the mature barnase coding region described above and a promoter region isolated from a seed coat-specific gene. A seed coat-specific gene is being isolated using the following steps. Seed coats were dissected from immature watermelon seeds and total cellular RNA was prepared using the guanidium isothiocyanate extraction procedure from Stratagene followed by LiCl precipitation as described by Ausubel et al. [Current Protocols in Molecular Biology 481-483 (1987)]. Total RNA of leaves was isolated using the method described by Baker et al. [Biotechniques 9(3), 268-272 (1990)]. Next a polyadenylated (polyA+) RNA fraction was prepared by two rounds of affinity chromatography on oligo(dT)-cellulose spin columns from Pharmacia following the manufacturer's procedure. A cDNA library made to this polyA+ RNA preparation was obtained from Stratagene. Duplicate filters made from plates of the library, or duplicate slot blots containing DNA made from each individual cDNA clone prepared according to Conkling et al. [Plant Physiol. 93, 1203-1211 (1990)] were differentially screened using <sup>32</sup>P-labeled cDNA probes made from the seed coat polyA+ RNA or from leaf polyA+ RNA according to Sargent [Guide to Molecular Cloning Techniques, Berger and Kimmel, eds. 423-432 (1987)]. Clones were

identified that hybridized only to the seed coat probe. The cloned cDNA sequences are verified as being derived from seed-specific RNAs by using them to probe a Northern blot, as described in BioRad's protocol, prepared with RNAs made from different plant tissues. A seed-specific cDNA is used to probe a genomic library made from watermelon DNA, made with a kit from NEN and using the included protocol, and the gene encoding the seed-specific RNA represented by the cDNA is identified. The genomic clone is mapped to locate the desired gene. The 5' end of the transcript is located by primer extension experiments as described in Ausubel et al. [Current Protocols in Molecular Biology 481-483 (1987)]. The promoter region including about 1 kilobase of sequence upstream of the transcription start site is prepared as a DNA fragment. This seed coat promoter is ligated to the barnase coding region followed by the Phaseolin 3' region, as described in the Example 13 above, creating the chimeric gene called SCP-bar-Ph 3'.

This chimeric seed coat disruption gene is made into an inactive form by adding a loxP-polyA-loxP DNA fragment between the promoter and coding region as described for the NOS/P-NptII gene in Example 5. The inactive SCP-lox-bar-Ph 3' gene is cloned into a binary vector between the T-DNA borders along with a chimeric kanamycin resistance selection marker gene. A diagram of the T-DNA of the binary vector is shown below:



This plasmid, called lox/SCPB/NptII, is transferred into a disarmed A. tumefaciens and the resulting strain is

used to obtain tobacco or Arabidopsis transformants as described previously.

Expression of the cre coding region is more effective either with the same developmentally controlled promoter or with the highly active 35S promoter. The cre coding region is placed under control of the same seed coat promoter used for the disruption gene making a chimeric SCP-cre-Ph 3' gene. This gene is cloned into a binary vector between the T-DNA borders along with a chimeric sulfonylurea resistance selection marker gene, creating the SCPCre/ALS plasmid. This plasmid is transferred into a disarmed A. tumefaciens and the resulting strain is used to obtain tobacco or Arabidopsis transformants as described previously.

Primary transformants containing lox/SCPB/NptII are crossed with primary transformants containing SCPCre/ALS, and with the homozygous Cre plants (containing the chimeric 35S-Cre gene) described in Section G and Example 8. Since the primary transformants are heterozygous, the seed produced from their crosses could carry none, both, or either one of the foreign DNA insertions. Therefore, progeny are screened for the presence of the two marker genes, ALS and NptII by germinating seeds on medium containing both chlorsulfuron and kanamycin. Resistant progeny that carry both the inactive seed coat disruption gene and the chimeric cre gene, are grown to maturity and selfed. Seed pods are checked for the absence of seed indicating that the seed coat disruption gene is activated by the loxP-cre system thereby disrupting production of F2 seed.

In crosses of lox/SCPB/NptII plants with the homozygous 35S-Cre plants, all progeny receive the chimeric Cre gene so only the kanamycin selection is necessary to identify progeny also containing the seed

coat disruption gene. Selected plants are grown to maturity and selfed. Seed pods are checked for the absence of seed indicating that the seed coat disruption gene is activated and effective in disrupting production of F2 seed.

Homozygous lines of lox/SCPB/NptII and of SCPCre/ALS plants are obtained as described previously and crossed. Seed pods are checked for seed production and seed viability is tested in germination assays. The presence of viable seed indicates that the inactive seed coat disruption gene maintains its inactive state in the seed coat of the developing F1 seed as predicted. Progeny are grown to maturity and selfed. Seed pods are checked for the absence of seed indicating that the seed coat disruption gene is activated and effective in disrupting production of F2 seed. Homozygous lox/SCPB/NptII plants are crossed with homozygous 35S-Cre plants and the production of viable F1 seed is tested. Progeny are grown to maturity, selfed, and the absence of F2 seed is observed.

It is to be appreciated that several modifications can be made to the subject invention described herein without departing from the spirit and scope thereof.

CLAIMS

What is Claimed is:

1. A method for producing site-specific  
5 recombination of DNA in plant cells, comprising:
  - i) introducing into the cells a first DNA  
sequence comprising a first lox site, and  
a second DNA sequence comprising a second  
lox site, and
  - 10 ii) contacting the lox sites with Cre, thereby  
producing the site-specific recombination.
2. A method as defined in Claim 1, wherein a third  
DNA sequence comprising a cre coding region is also  
15 introduced into the cells.
3. A method as defined in Claim 2, wherein the  
third DNA sequence comprises a promoter that is active  
in plant cells and expression of the cre gene is  
20 produced by direction of the promoter.
4. A method as defined in Claim 3, wherein the  
first and second DNA sequences are introduced into the  
cells connected by a pre-selected DNA segment.  
25
5. A method as defined in Claim 4, wherein the  
first and second lox sites have the same orientation and  
the site specific recombination of DNA is a deletion of  
the pre-selected DNA segment.  
30
6. A method as defined in Claim 5, wherein the cre  
coding region is derived from bacteriophage P1.

7. A method as defined in Claim 5, wherein the first and second lox sites are loxP or derivatives thereof.

5           8. A method as defined in Claim 5, wherein the first and second lox sites are loxP.

9. A method as defined in Claim 5, wherein the pre-selected DNA segment is selected from the group  
10 consisting of a gene, a coding region, and a DNA sequence that influences gene expression in plant cells.

10. A method as defined in Claim 5, wherein the pre-selected DNA segment is an undesired marker or trait  
15 gene.

11. A method as defined in Claim 4, wherein the first and second lox sites have opposite orientations and the site-specific recombination is an inversion of  
20 the nucleotide sequence of the pre-selected DNA segment.

12. A method as defined in Claim 11, wherein the cre coding region is derived from bacteriophage P1.

25           13. A method as defined in Claim 12, wherein the first and second lox sites are loxP or derivatives thereof.

14. A method as defined in Claim 12, wherein the  
30 first and second lox sites are loxP.

15. A method as defined in Claim 11, wherein the pre-selected DNA segment is selected from the group consisting of a gene, a coding region, and a DNA  
35 sequence that influences gene expression in plant cells.

16. A method as defined in Claim 1, wherein the first and second DNA sequences are introduced into two different DNA molecules and the site-specific  
5 recombination is a reciprocal exchange of DNA segments proximate to the lox sites.

17. A method as defined in Claim 16, wherein the cre  
coding region is derived from bacteriophage P1.  
10

18. A method as defined in Claim 17, wherein the first and second lox sites are loxP and derivatives thereof.

19. A method as defined in Claim 18, wherein the first and second lox sites are loxP.  
15

20. A method of excising exogenous genes or DNA segments in transgenic plants, comprising:  
20 i) introducing into the cells a DNA sequence comprising a first lox site, a second lox site in the same orientation as the first lox site, and a gene or a DNA sequence therebetween, and  
25 ii) contacting the lox sites with Cre, thereby excising the heterologous gene or DNA sequence.

21. A method as defined in Claim 20, wherein the  
30 gene is an undesired marker or trait gene.

22. A plant cell transformed with a DNA sequence comprising at least one lox site.

23. A plant cell containing Cre protein.  
35

24. A plant cell transformed with a cre coding region.

5        25. A plant containing cells transformed with a DNA sequence comprising at least one lox site.

26. A plant of Claim 25 wherein the DNA sequence comprises two lox sites and a cre coding region.

10

27. A plant as defined in Claim 25, and having agronomic or horticultural utility.

15        28. A plant containing cells transformed with a cre coding region.

29. A plant as defined in Claim 28, and having agronomic or horticultural utility.

20        30. A plasmid having at least one lox site and a pre-selected DNA segment selected from the group consisting of a gene, a coding region, and a DNA sequence that influences gene expression in plant cells.

25        31. A plasmid as defined in Claim 30, wherein the DNA sequence is a polyadenylation nucleotide sequence derived from the Rubisco small subunit gene.

30        32. A plasmid as defined in Claim 30, wherein the DNA sequence is a selection marker.

33. A plasmid as defined in Claim 30, wherein the DNA sequence is a promoter.



34. A plasmid as defined in Claim 30, wherein the DNA sequence is a regulatory nucleotide sequence.

35. A plasmid having a cre coding region and a  
5 promoter that is active in plant cells.

36. A DNA sequence comprising at least one lox site and a pre-selected DNA segment selected from the group consisting of a gene, a coding region, and a DNA  
10 sequence that influences gene expression in plant cells.

37. A DNA sequence comprising a cre coding region and a promoter that is active in plant cells.

15 38. Plasmid Cre/Hpt-A characterized by the restriction enzyme map shown in Figure 1A, or a derivative thereof.

39 Plasmid Cre/Hpt-B characterized by the  
20 restriction enzyme map shown in Figure 1B, or a derivative thereof.

40. Plasmid loxP/NptII/Hra characterized by the restriction enzyme map shown in Figure 1C, or a  
25 derivative thereof.

41. Plasmid pZ4loxAG characterized by the restriction enzyme map shown in Figure 4, or a  
30 derivative thereof.

42. The method of Claim 1 useful in the manufacture of seedless produce.

1/7

FIG. 1A

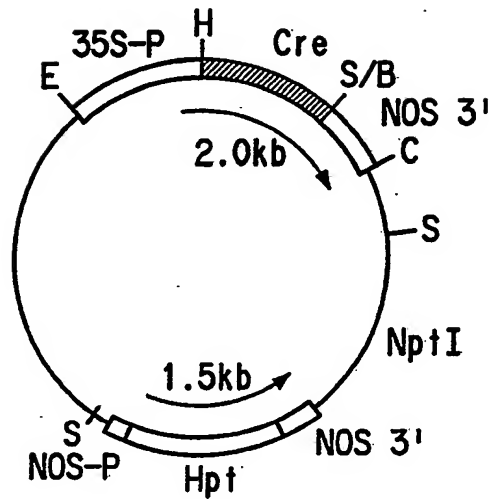
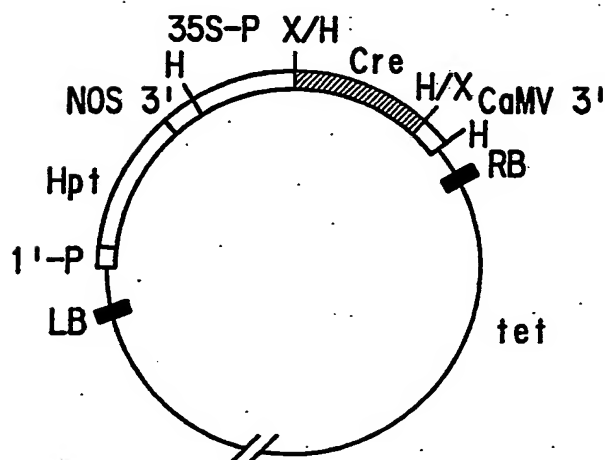


FIG. 1B



2/7

FIG. 1B

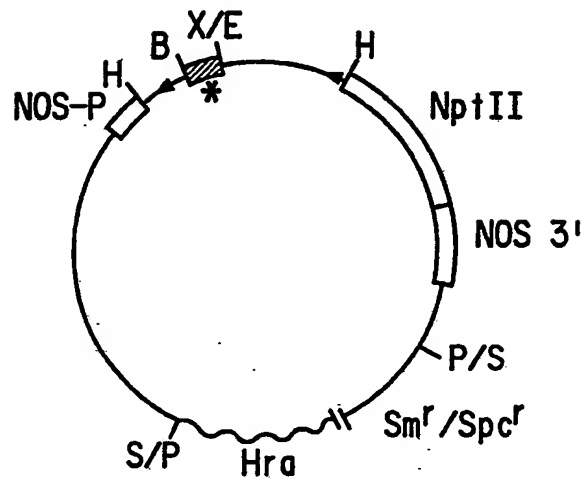
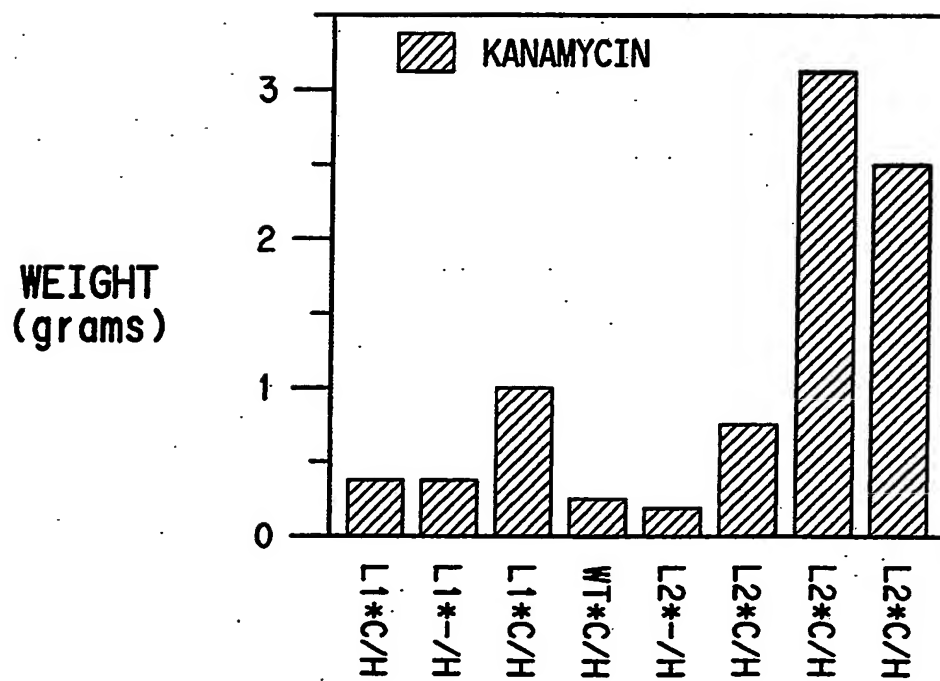
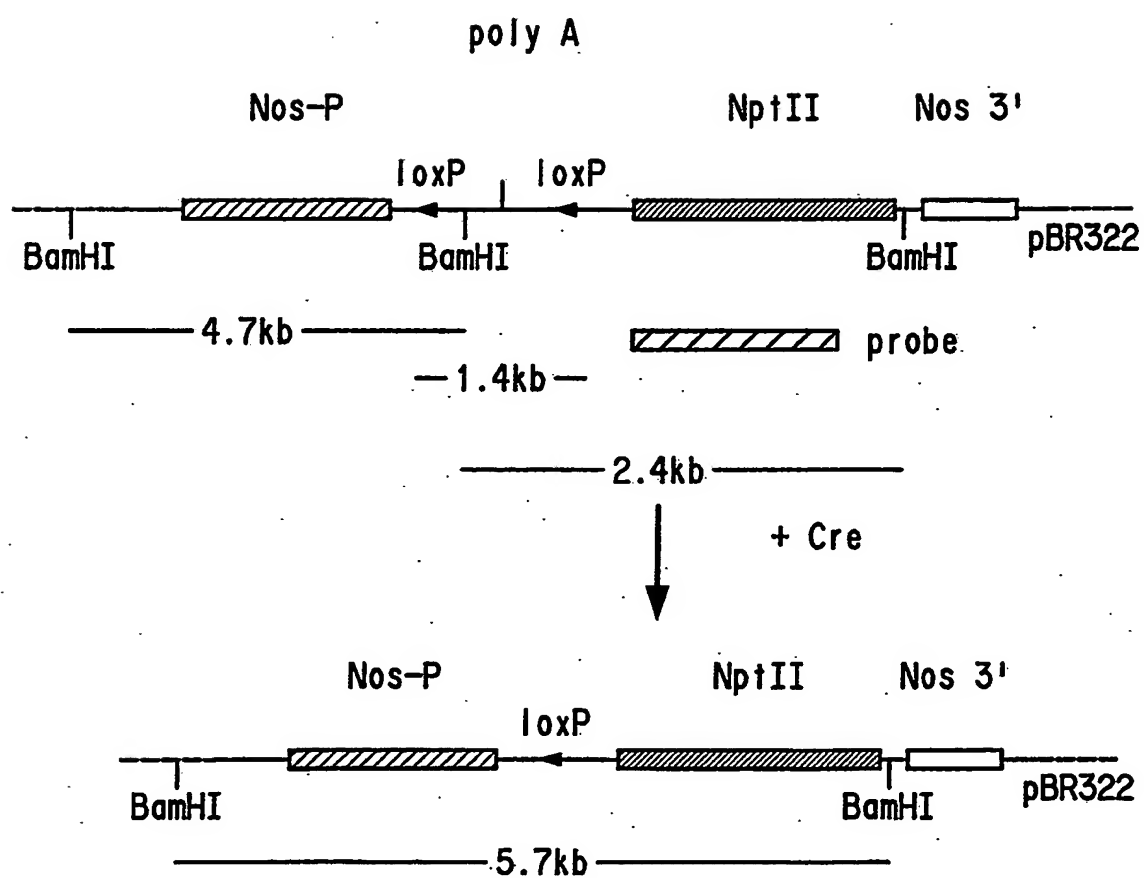


FIG. 2A



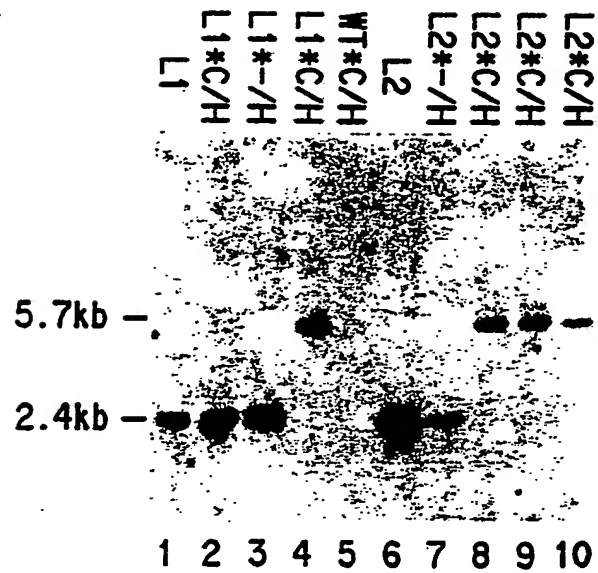
3/7

FIG. 2B



4/7

FIG.2C





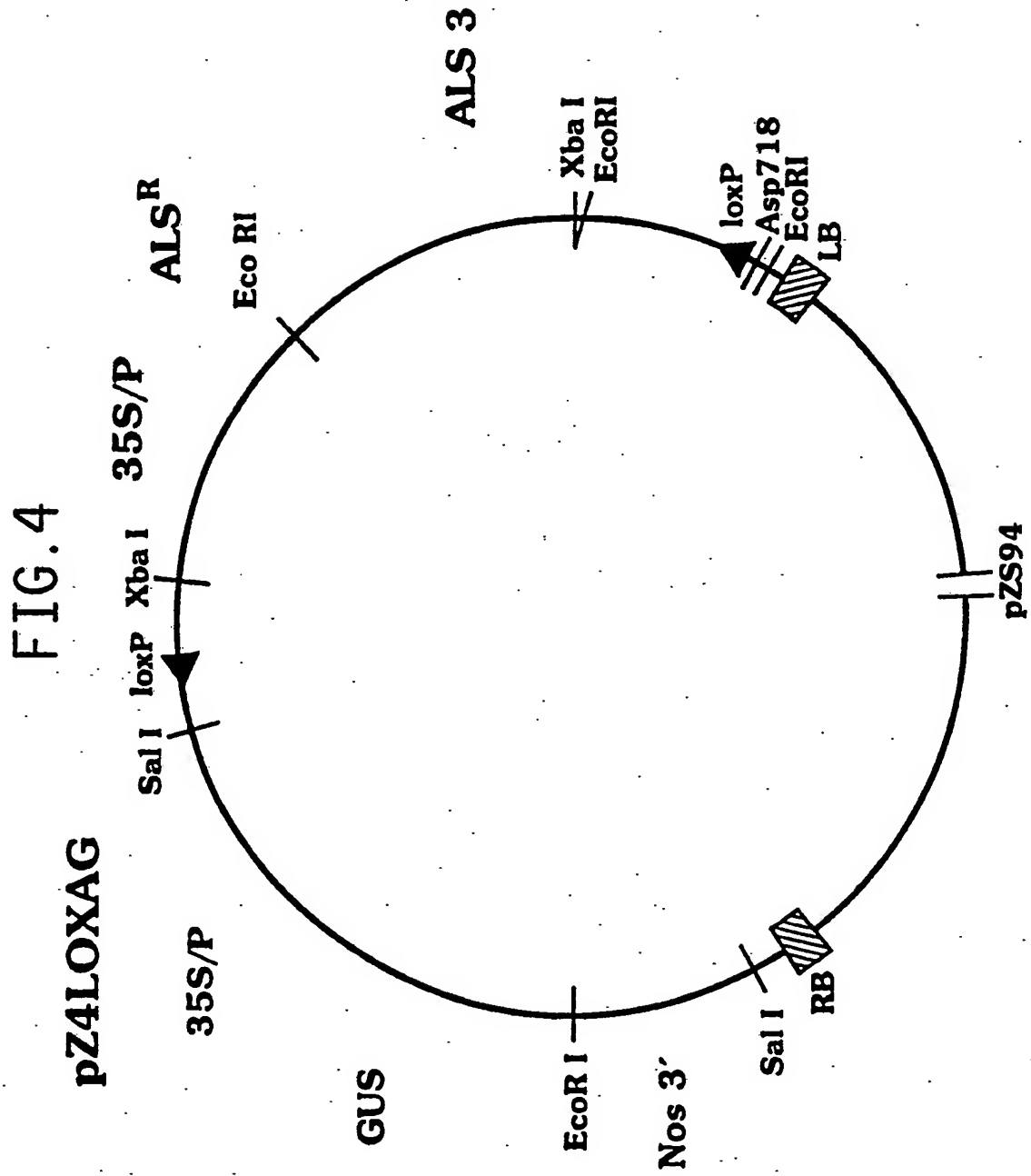
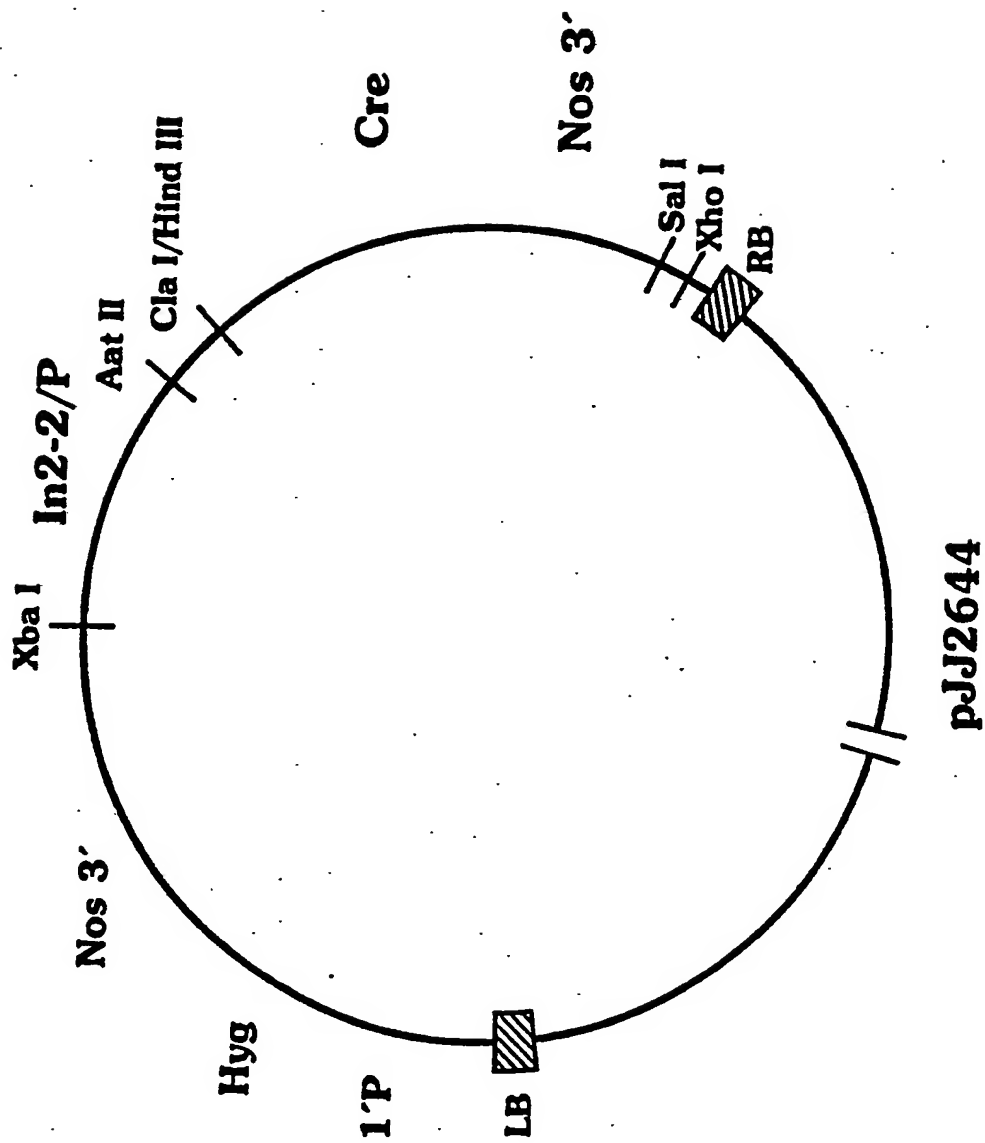


FIG. 5

pBSCre103





# INTERNATIONAL SEARCH REPORT

International Application No PCT/US 90/07295

<b>I. CLASSIFICATION OF SUBJECT MATTER</b> (If several classification symbols apply, indicate all) *		
According to International Patent Classification (IPC) or to both National Classification and IPC		
IPC <sup>5</sup> : C 12 N 15/82, C 12 N 5/10, A 01 H 5/00		
<b>II. FIELDS SEARCHED</b>		
Minimum Documentation Searched <sup>7</sup>		
Classification System	Classification Symbols	
IPC <sup>5</sup>	C 12 N, A 01 H	
Documentation Searched other than Minimum Documentation to the extent that such Documents are included in the Fields Searched <sup>8</sup>		
<b>III. DOCUMENTS CONSIDERED TO BE RELEVANT</b> <sup>9</sup>		
Category <sup>9</sup>	Citation of Document, <sup>11</sup> with indication, where appropriate, of the relevant passages <sup>12</sup>	Relevant to Claim No. <sup>13</sup>
X	Nucleic Acids Research, vol. 17, no. 1, 1989, B. Sauer et al.: "Cre-stimulated recombination at loxP-containing DNA sequences placed into the mammalian genome", pages 147-161 see abstract; page 148, "Plasmid constructions"; (cited in the application)	30, 32-34, 36
A	--	1-29, 31, 35, 37
A	EP, A, 0220009 (DU PONT) 29 April 1987 see the whole document	1-37
	--	
P, X	Mol. Gen. Genet, vol. 223, September 1990, Springer-Verlag 1990, J. Odell et al.: "Site-directed recombination in the genome of transgenic tobacco", pages 369-378 see the whole article ./.	1-10, 20-40
<div style="display: flex; justify-content: space-between;"> <div style="width: 45%;"> <p>* Special categories of cited documents: <sup>10</sup></p> <p>"A" document defining the general state of the art which is not considered to be of particular relevance</p> <p>"E" earlier document but published on or after the international filing date</p> <p>"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)</p> <p>"O" document referring to an oral disclosure, use, exhibition or other means</p> <p>"P" document published prior to the international filing date but later than the priority date claimed</p> </div> <div style="width: 45%;"> <p>"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention</p> <p>"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step</p> <p>"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.</p> <p>"A" document member of the same patent family</p> </div> </div>		
<b>IV. CERTIFICATION</b>		
Date of the Actual Completion of the International Search	Date of Mailing of this International Search Report	
26th March 1991	23. 04. 91	
International Searching Authority	Signature of Authorized Officer	
EUROPEAN PATENT OFFICE	<div style="display: flex; align-items: center;"> <div style="border: 1px solid black; padding: 2px 5px; margin-right: 10px;">M. PEIS</div> <div style="font-family: cursive; font-size: 1.2em;">M. Peis</div> </div>	

III. DOCUMENTS CONSIDERED TO BE RELEVANT (CONTINUED FROM THE SECOND SHEET)		
Category *	Citation of Document, ** with indication, where appropriate, of the relevant passages	Relevant to Claim No.
P,X	Gene, vol. 91, 1990, Elsevier Science Publishers B.V., E.C. Dale et al.: "Intra- and inter- molecular site-specific recombination in plant cells mediated by bacterio- phage P1 recombinase", pages 79-85 see the whole article  -----	11-19

US 9007295 .  
SA 43346

The European Patent Office is in no way liable for these particulars which are merely given for the purpose of information.

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
EP-A- 0220009	29-04-87	JP-A- 62175184 US-A- 4959317	31-07-87 25-09-90
-----			



## INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

<b>(51) International Patent Classification <sup>5</sup> :</b> <b>C12N 15/00, C12P 19/34, 21/06</b> <b>C07H 21/00</b>	<b>A1</b>	<b>(11) International Publication Number:</b> <b>WO 91/16427</b> <b>(43) International Publication Date:</b> 31 October 1991 (31.10.91)
<b>(21) International Application Number:</b> PCT/US91/02910 <b>(22) International Filing Date:</b> 24 April 1991 (24.04.91) <b>(30) Priority data:</b> 513,957 24 April 1990 (24.04.90) US <b>(71) Applicant:</b> STRATAGENE [US/US]; 11099 North Torrey Pines Road, La Jolla, CA 92037 (US). <b>(72) Inventors:</b> SHORT, Jay, M. ; 320 Delage Drive, Encinitas, CA 92024 (US). SORGE, Joseph, A. ; 17021 Circa Del Sur, Rancho Santa Fe, CA 92067 (US). <b>(74) Agents:</b> BIGGS, Suzanne, L. et al.; 611 West Sixth Street, 34th Floor, Los Angeles, CA 90017 (US).		<b>(81) Designated States:</b> AT (European patent), AU, BE (European patent), CA, CH (European patent), DE (European patent), DK (European patent), ES (European patent), FI, FR (European patent), GB (European patent), GR (European patent), IT (European patent), JP, KR, LU (European patent), NL (European patent), NO, SE (European patent).  <b>Published</b> <i>With international search report.</i>
<b>(54) Title:</b> METHODS FOR PHENOTYPE CREATION FROM MULTIPLE GENE POPULATIONS  <b>(57) Abstract</b>  Methods of producing biological agents which express a desired identifiable phenotype are provided. These methods include bringing together populations of diverse replicas of nucleotide sequences to give a plurality of combined nucleotide sequences, each comprising one member of each population, expressing the combined nucleotide sequences to give a phenotype and identifying those biological agents expressing the desired phenotype.		

**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	ES	Spain	MG	Madagascar
AU	Australia	FI	Finland	ML	Mali
BB	Barbados	FR	France	MN	Mongolia
BE	Belgium	GA	Gabon	MR	Mauritania
BF	Burkina Faso	GB	United Kingdom	MW	Malawi
BG	Bulgaria	GN	Guinea	NL	Netherlands
BJ	Benin	GR	Greece	NO	Norway
BR	Brazil	HU	Hungary	PL	Poland
CA	Canada	IT	Italy	RO	Romania
CF	Central African Republic	JP	Japan	SD	Sudan
CG	Congo	KP	Democratic People's Republic of Korea	SE	Sweden
CH	Switzerland	KR	Republic of Korea	SN	Senegal
CI	Côte d'Ivoire	LI	Liechtenstein	SU	Soviet Union
CM	Cameroon	LK	Sri Lanka	TD	Chad
CS	Czechoslovakia	LU	Luxembourg	TG	Togo
DE	Germany	MC	Monaco	US	United States of America
DK	Denmark				

DESCRIPTIONMethods for Phenotype Creation  
From Multiple Gene PopulationsCross Reference to Related Application

This is a continuation-in-part application of copending application Serial No. 513,957, filed April 24, 1990 which is a continuation-in-part of Serial No. 353,235, 5 filed May 16, 1989, and Serial No. 353,241, filed May 17, 1989, the disclosures of which are hereby incorporated by reference.

Field of the Invention

The present invention relates to methods for randomly 10 combining populations of nucleotide sequences and selecting those combinations coding for a desired predetermined phenotype.

Background of the Invention

The production of genetic variants, including variants 15 of both polypeptides and organisms such as bacteria and phage, has been a goal in the work of many individuals involved in recombinant DNA technologies. For example, researchers have beneficially relied upon random genetic recombination in the past for the production of new and 20 useful microorganisms. Genetic recombination includes a variety of processes that produce new linkage relationships of genes or parts of genes. Genetic recombination is often subdivided into general genetic recombination, which takes place between homologous chromosomes, more or 25 less anywhere along their length, and recombination that does not require extensive homology. The latter category includes site-specific recombination, which depends upon the existence of specific sites in one or more molecules and which includes interactions of viral genomes and 30 insertion sequences with chromosomes of prokaryotes and

eukaryotes, and less well defined instances of recombination that appear to require neither extensive homology nor special sites. Variable gene expression can also result in production of various combinations of polypeptides, the  
5 immune system being one example of such protein combination.

The immune system of a mammal is one of the most versatile biological systems; probably greater than  $1.0 \times 10^7$  antibody specificities can be produced. Indeed, a  
10 great deal of contemporary biological and medical research is directed toward tapping this repertoire. During the last decade, furthermore, there has been a dramatic increase in the ability to harness the output of the immune system. The development of the hybridoma method-  
15 ology by Kohler and Milstein has made it possible to produce monoclonal antibodies, i.e., a composition of antibody molecules of single epitope specificity, from the repertoire of antibodies induced during an immune response. Monoclonal antibodies have been generated in  
20 the past from hybridomas, generated by fusing antibody-secreting lymphocytes with an immortal cell line, such as myeloma.

Although standard hybridoma technology has been extremely valuable, the screening of fused cells to iden-  
25 tify hybridomas expressing useful antibody molecules is labor intensive, time consuming and expensive. Moreover, the standard technology yields rodent antibody molecules that have two clear disadvantages. The first is that subtle variations in certain human antigenic systems, such  
30 as major histocompatibility proteins, are not easily distinguished by non-primate antibodies. Therefore, rodent antibodies may not provide the repertoire of specificities needed to distinguish certain polymorphic antigenic determinants. In other words, current methods for generating  
35 monoclonal antibodies are not capable of efficiently surveying the entire antibody response induced by a particular immunogen. Thus, in an individual animal there

**SUBSTITUTE SHEET**

are at least 5-10,000 different B-cell clones capable of generating unique antibodies to a small relatively rigid immunogens, such as, for example dinitrophenol. Further, because of the process of somatic mutation during the generation of antibody diversity, essentially an unlimited number of unique antibody molecules may be generated. In contrast to this vast potential for different antibodies, current hybridoma methodologies typically yield only a few hundred different monoclonal antibodies per fusion. A second major drawback in hybridoma technology is that rodent antibodies are highly immunogenic in humans, and can preclude their continued use in patients for diagnostic or therapeutic purposes.

One alternative is to produce human cells that express antibody. Unfortunately, it is quite difficult to identify and produce pure human monoclonal antibodies. Standard methods used to immortalize antibody-producing cells are less than satisfactory. One approach that circumvents the need for human hybridoma cells has been to use recombinant DNA technology to express fusion antibody proteins. These molecules have amino terminal variable domains of the light and heavy chains derived from a specific rodent monoclonal antibody and the carboxy terminal constant region domains derived from a human antibody. The use of human constant regions diminishes the human anti-globulin immune response, avoiding the stimulation of anti-isotypic antibody-producing B cells. However, the rodent-derived variable region framework domains still elicit a response that is more severe than a variable domain response directed against a pure human antibody.

In an effort to avoid the anti-idiotypic response directed against the rodent framework regions of the domains, some researchers have taken a human antibody and replaced the hypervariable regions (CDRs) with hypervariable regions from a rodent antibody specific for a selected antigen. Although such antibodies may have an



affinity for antigen comparable to the parent rodent antibody, the process of grafting all rodent CDRs into a human immunoglobulin gene is technically challenging.

Aside from repertoire specificity and immunogenicity,  
5 other drawbacks in producing monoclonal antibodies with the hybridoma methodology include genetic instability and low production capacity of hybridoma cultures. One means by which the art has attempted to overcome these latter two problems has been to clone the immunoglobulin-  
10 producing genes from a particular hybridoma of interest into a procaryotic expression system. See, for example, Robinson et al., PCT Publication No. WO 89/0099; Winter et al., European Patent Publication No. 0239400; Reading, U.S. Patent No. 4,714,681; and Cabilly et al., European  
15 Patent Publication No. 0125023.

The immunologic repertoire of vertebrates has recently been found to contain genes coding for immunoglobulins having catalytic activity. Tramontano et al., Sci., 234:1566-1570 (1986); Pollack et al., Sci.,  
20 234:1570-1573 (1986); Janda et al., Sci., 244:437-440 (1989). The presence of, or the ability to induce the repertoire to produce, antibody molecules capable of a catalyzing chemical reaction, i.e., acting like enzymes, had previously been postulated almost 20 years ago by W.  
25 P. Jencks in Catalysis in Chemistry and Enzymology, McGraw-Hill, N.Y. (1969).

It is believed that one reason the art failed to isolate catalytic antibodies from the immunological repertoire earlier, and its failure to isolate many to  
30 date even after their actual discovery, is the inability to screen a large portion of the repertoire for the desired activity. Another reason is believed to be the bias of currently available screening techniques, such as the hybridoma technique, towards the production high  
35 affinity antibodies inherently designed for participation in the process of neutralization, as opposed to catalysis.

In an attempt to enhance the designed recombination of desired DNA sequences or the desired combination of otherwise randomly generated polypeptides, including the identification and production of pure human monoclonal antibodies, we have pursued alternative approaches for the production and screening of such nucleotide sequences and polypeptides.

#### Summary of the Invention

The present invention is directed to methods for producing biological agents having a desired novel phenotype wherein this phenotype results from expression of a particular combined nucleotide sequence and wherein said phenotype can be used to identify the biological agents having the particular combined nucleotide sequence and distinguish them from biological agents having other combined nucleotide sequences. The desired phenotype is typically a phenotype which is not normally expressed by the parent nucleotide sequences. In one embodiment these methods comprise first replicating at least portions of two parent nucleotide sequences. The replicating step yields a population of diverse replicas of parent nucleotide sequences. In one embodiment, each parent nucleotide sequence initially comprises a population (or family) of diverse nucleotide sequences which is replicated to give a population of diverse replicas. Alternatively, a population of diverse replicas is generated by replicating a parent nucleotide sequence under conditions which allow mutations to occur which generates diversity from one parent nucleotide sequence and results in a population of diverse replicas. In one aspect, the parent nucleotide sequences may comprise a single DNA molecule or alternatively the parent nucleotide sequences comprise separate DNA molecules. Where the parent nucleotide sequences comprise one DNA molecule, after replication, the resulting populations of diverse replicas derived from each parent nucleotide sequence are separated. The populations of

diverse replicas are then brought together, preferably in a random manner, to produce combined nucleotide sequences wherein each combined nucleotide sequence comprises one member of each population of diverse replicas. The parent  
5 nucleotide sequences may be suitably replicated and brought together according to the various methods described herein for replication and recombination of nucleotide sequences and generation of combinatorial libraries. The combined nucleotide sequences are  
10 expressed in biological agents. Such biological agents may comprise a host cell, or alternatively, a plasmid, bacteriophage or virus, or nucleic acid vector, and such suitable means for expression are described herein. In one embodiment< expression may constitute the mere exist-  
15 ence of the nucleotide sequences in the same biological agent. Then, the biological agents which express the desired phenotype are identified. If desired, the phenotype is used to distinguish those biological agents expressing the particular combined nucleotide sequence  
20 from biological agents expressing other combined nucleotide sequences. The desired phenotype may comprise a polypeptide, more than one polypeptide, or a multimeric polypeptide, the expression of which is detectable. Alternatively, the phenotype may comprise synthesis of one  
25 or more RNA molecules. Optionally, either the polypeptides or RNA may exhibit enzymatic activity or receptor activity; or the DNA or RNA may simply act as a target for interaction with other molecules.

The present invention provides novel methods for the  
30 cloning of cells having novel phenotypes. These methods generally include the use of a combinatorial library selection system to generate a diverse collection of clones. In one aspect, the methods utilize at least two starting populations of nucleotide sequences which can be  
35 recombined to form a library of clones containing nucleotide sequences from each of the parent populations. These methods can be utilized, therefore, to create cells having

novel phenotypes, that is, cells having a new and desired combination of expressed polypeptides. These methods can also be used for the production of new combinations of polypeptides, including the polypeptides utilized in the formation of biologically competent immunoglobulin molecules. In accordance with the latter object of the invention, these methods can be used to screen a larger portion of the immunological repertoire for receptors having a preselected activity than has heretofore been possible, thereby overcoming the before-mentioned inadequacies of the hybridoma technique.

In another embodiment, the present invention contemplates a gene library comprising an isolated admixture of at least about  $10^3$ , preferably at least about  $10^4$  and more preferably at least  $10^5$   $V_H$ -and/or  $V_L$ -coding DNA homologs, a plurality of which share a conserved antigenic determinant. Preferably, the homologs are present in a medium suitable for in vitro manipulation, such as water, phosphate buffered saline and the like, which maintains the biological activity of the homologs.

In one embodiment, at least two starting populations of DNA sequence-containing vectors are physically combined by any of several techniques, including those described herein, to form a library of clones containing DNA sequences from each of the parent populations. Alternatively there may be more than two gene families and the vectors produced thereby may contain a random assortment of one member of each gene family to create the identifiable characteristic. These vectors can then be transferred to desired host cells to create in vivo novel combinations of phenotypic characteristics in the host cell. Methods of combining desired DNA sequences include the use of restriction digestion and ligation, homologous recombination, and site-specific recombination by methods including intergrase-related proteins, flp recombinase-catalyzed recombination, the cre-lox system of bacteriophage P1, and the use of transposons.

In a still further embodiment, the present invention contemplates vectors for use in the methods which comprise, in addition to random DNA sequences from the starting gene family populations, DNA sequences which  
5 facilitate the region-specific, random recombination together of at least one gene from each starting gene family population. Sequences enabling the recombination of these vectors include the use of functional flp recombination sequences, functional loxp recombination  
10 sequences, at sequences recognized by integrase-related proteins from lambdoid bacteriophages, and terminal repeat sequences recognized by transposases. Thus, the present invention also includes methods for the combinatorial generation of phenotypes, including a method of producing  
15 a nucleic acid vector encoding two or more desired genes each from a family of genes, said genes being capable of producing a characteristic that can be used to identify the vector encoding said genes from other vectors encoding other members of the families of genes, which method  
20 comprises:

a) randomly inserting into vectors one member of a first family of genes and one member from one or more other families of genes so that a population of vectors are created wherein each vector may contain one of the  
25 genes from said first gene family and one of the genes from each of said other gene families;

b) identifying within said population of vectors a vector capable of detectably producing a desired characteristic resulting from the inclusion of one gene from  
30 said first gene family and one gene from each of said other gene families, and using said characteristic to distinguish the vector from other vectors within the population containing undesired combinations of gene members from said gene families.

35 Suitable vectors for use according to the methods of the present invention include plasmid or cosmid vectors or, alternatively, phage vectors. Suitable host cells for

expressing the vectors comprise either eukaryotic cells or prokaryotic cells. Preferred eukaryotic cells include mammalian cells. In one preferred aspect, the vectors comprise lambda bacteriophage and host cells comprise E.

5 Coli.

Preferably, the genes are combined in vivo.

Various suitable methods may be used for the identification of a particular vector within the recombinant vector population. These methods include (a) the interaction of sequence-specific nucleic acids with genes from  
10 the individual families which were combined: (b) the hybridization of nucleic acid probes with genes from the gene families; (c) the expression of one or both genes from the gene families as an RNA molecule; and (d) the  
15 expression of one or both genes as an identifiable protein molecule. Optionally, such an identifiable protein molecule may contain a binding site for another molecule, an epitope recognized by an antibody, or an immune molecule binding site for an epitope. In a preferred  
20 identification method, both genes express an RNA and/or polypeptide and said RNAs and/or polypeptides physically interact with a host to create an identifiable characteristic. Both genes may express polypeptides that physically interact to form a neo-epitope recognized by an  
25 immune molecule or polypeptides that physically interact to form a binding site for another molecule. Optionally those polypeptides are derived from antibody genes such that the interaction of both polypeptides forms an antigen binding site.

30 In another preferred aspect, the vectors produced according to the present invention contain a single promoter that expresses the genes from the gene families. Alternatively, the genes from the gene families are each expressed from their own promoter.

35 In a still further embodiment, the present invention contemplates the creation of combinations of two or more nucleotide sequence families (or populations) by in vitro

recombination. Such in vitro recombination could be carried out using specific recombination target sequences and specific recombinases (like flp recombinase), or by using homologous sequences shared by both nucleotide  
5 sequence populations to facilitate homologous recombination.

One method to accomplish a form of homologous recombination in vitro is by using in vitro nucleic acid amplification methods such as the polymerase chain reaction (PCR). If both of two populations of DNA sequences share a region of homology, then it is possible during the PCR for base-pairing to occur between single stranded nucleic acid molecules from both populations of nucleotide sequences. If such base pairing creates a "primer-  
15 template complex" that can be used by a polymerase to begin synthesis of complementary strands, then a fusion product is created which will contain sequences from both nucleotide sequence populations (See Figure 21 here). If the shared region of homology is present on most or all of  
20 the two nucleotide sequence populations, then most or all of the nucleotide sequences can participate in such recombination. Thus, a combinatorial population of fusion nucleotide sequences can be produced, and subsequently inserted into a single expression vector for expression of  
25 the nucleotide sequence from both sequence families. Such a combinatorial population of expressed sequences can then be screened for new phenotypes that would not be present if the sequences from only one population of nucleotide sequences were expressed, and would be present only with  
30 expression of particular combinations comprising a nucleotide sequence from each population. For example, such phenotypes could comprise the creation of heterodimeric proteins where one subunit of the dimer is encoded by one nucleotide sequence family and the other subunit of the  
35 dimer is encoded by the other nucleotide sequence family.

Thus, the present invention is directed to methods of creating diversity, namely populations of diverse replicas

of nucleotide sequences which may be combined to give a diversity of phenotypes, from which a desired phenotype may be selected. Such diversity may be generated starting with a single DNA molecule which is treated to create  
5 diversity, such as by mutagenesis or by starting with a family of nucleotide sequences (or genes) or a combinatorial library.

For example, one may start with a plasmid containing antibody sequences coding for both a light chain and a  
10 heavy chain which has been isolated from a known monoclonal-antibody producing cell line. The nucleotide sequences coding for the light chain and the heavy chain may be individually amplified (using a method such as PCR) under conditions that mutated sequences are generated to  
15 create a population of mutated sequences. The individual populations of mutated sequences may be used to make combinatorial libraries which are then used to create novel phenotypes. Alternatively, these individual populations of mutated sequences may be combined using techniques such  
20 as fusion polynucleotide amplification (for example) fusion PCR (as described herein) and used to generate novel phenotypes. These novel phenotypes may include antibodies having enhanced antigen binding characteristics.

25 According to another aspect of the present invention, one or more genetically distinct phage may be lytically replicated, conditions which are somewhat mutagenic, to generate a population(s) of diverse phage. Phage having phenotypes distinct from the originals may be generated by  
30 cleavage such as by a restriction endonuclease, followed by mixing of phage populations, and ligation, followed by selection for expression of desired phenotypes. In this way phage having diverse phenotypes distinct from the parental phage may be generated combinatorially.

35 In another embodiment, the methods are utilized to produce novel human antibody-expressing DNA sequences. First, an immunoglobulin heavy chain variable region  $V_H$



gene library containing a substantial portion of the  $V_H$  gene repertoire of a vertebrate is synthesized. In preferred embodiments, the  $V_H$ -coding gene library contains at least about  $10^3$  and more preferably at least about  $10^4$  and  
5 more preferably at least about  $10^5$  different  $V_H$ -coding nucleic acid strands referred to herein as  $V_H$ -coding DNA homologs.

The gene library can be synthesized by various methods, depending on the starting material. Where the  
10 starting material is a plurality of  $V_H$ -coding genes, the repertoire is subjected to two distinct primer extension reactions. The first primer extension reaction uses a first polynucleotide synthesis primer capable of initiating the first reaction by hybridizing to a nucleotide  
15 sequence conserved (shared by a plurality of genes) within the repertoire. The first primer extension reaction produces a plurality of different  $V_H$ -coding homolog complements (nucleic acid strands complementary to the genes in the repertoire). The second primer extension  
20 reaction produces, using the complements as templates, a plurality of different  $V_H$ -coding DNA homologs. The second primer extension reaction uses a second polynucleotide synthesis primer that is capable to initiating the second reaction by hybridizing to a nucleotide sequence conserved  
25 among a plurality of  $V_H$ -coding gene complements.

Where the starting material is a plurality of complements of different  $V_H$ -coding genes provided by a method other than the first primer extension reaction, the repertoire is subjected to the above-discussed second  
30 primer extension reaction. That is, where the starting material is a plurality of different  $V_H$ -coding gene complements produced by a method such as denaturation of double strand genomic DNA, chemical synthesis and the like, the complements are subjected to a primer extension  
35 reaction using a polynucleotide synthesis primer that hybridizes to a plurality of the different  $V_H$ -coding gene complements provided. Of course, if both a repertoire of

V-coding genes and their complements are present in the starting material, both approaches can be used in combination.

A  $V_H$ -coding DNA homolog, i.e., a gene coding for a  
5 receptor capable of binding the preselected ligand, is then segregated from the library to produce the isolated gene. This may be accomplished by operatively linking for expression a plurality of the different  $V_H$ -coding DNA homologs of the library to an expression vector. The  $V_H$ -  
10 expression vectors so produced are introduced into a population of compatible host cells, i.e., cells capable to expressing a gene operatively linked for expression to the vector. The transformants are cultured under conditions for expressing the receptor coded for by the  $V_H$ -coding DNA  
15 homolog. The transformants are cloned and the clones are screened for expression of a receptor that binds the preselected ligand. Any of the suitable methods well known in the art for detecting the binding of a ligand to a receptor can be used. A transformant expressing the  
20 desired activity is then segregated from the population to produce the isolated gene.

A receptor having a preselected activity produced by a method of the present invention, preferably a  $V_H$  or  $F_v$  as described herein, is also contemplated.

25 The present invention also encompasses products produced by the methods of the invention, such as the biological agents produced thereby, also the expression products of these methods such as polypeptides and nucleic acids, vectors produced and kits comprising any of the  
30 products of the claimed methods.

#### Brief Description of the Drawings

In the drawings forming a portion of this disclosure:

Figure 1 illustrates a schematic diagram of the immunoglobulin molecule showing the principal structural  
35 features. The circled area on the heavy chain represents the variable region ( $V_H$ ), a polypeptide containing a

biologically active (ligand binding) portion of that region, and a gene coding for that polypeptide, are produced by the methods of the present invention. Sequences L03, L35, L47 and L48 could not be classified  
5 into any predefined subgroups.

Figure 2A is a diagrammatic sketch of an H chain of human IgG (IgG1 subclass). Numbering is from the N-terminus on the left to the C-terminus on the right. Note the presence of four domains, each containing an intra-  
10 chain disulfide bond (S-S) spanning approximately 60 amino acid residues. The symbol CHO stands for carbohydrate. The V region of the heavy (H) chain ( $V_H$ ) resembles  $V_L$  in having three hypervariable CDR (not shown).

Figure 2B is a diagrammatic sketch of a human K chain (Panel 1). Numbering is from the N-terminus on the left to the C-terminus on the right. Note the intrachain disulfide bond (S-S) spanning about the same number of amino acid residues in the  $V_L$  and  $C_L$  domains. Panel 2 shows the locations of the complementarily-determining  
20 regions (CDR) in the  $V_L$  domain. Segments outside the CDR are the framework segments (FR).

Figure 3 depicts the amino acid sequence of the  $V_H$  regions of 19 mouse monoclonal antibodies with specificity for phosphorylcholine. The designation HP indicates that  
25 the protein is the product of a hybridoma. The remainder are myeloma proteins. (From Gearhart et al., Nature, 291:29, 1981.)

Figure 4 illustrates the results obtained from PCR amplification of mRNA obtained from the spleen of a mouse  
30 immunized with FITC. Lanes R17-R24 correspond to amplification reactions with the unique 5' primers (2-9, Table 1) and the 3' primer (12, Table 1), R16 represents the PCR reaction with the 5' primer containing inosine (10, Table 1) and 3' primer (12, Table 1). Z and R9 are the ampli-  
35 fication controls; control Z involves the amplification of  $V_H$  for a plasmid (PLR2) and R9 represents the amplification

from the constant regions of spleen mRNA using primers 11 and 13 (Table 1).

Figure 5 depicts nucleotide sequences of clones from the cDNA library of the PCR amplified  $V_H$  regions in Lambda ZAP vector. The N-terminal 110 bases are listed here and the underlined nucleotides represent CDR1 (complementary determining region).

Figures 6A and 6B depict the sequence of the synthetic DNA insert inserted into Lambda ZAP vector to produce Lambda Zap II  $V_H$  (6A) and Lambda Zap  $V_L$  (6B) expression vectors. The various features required for this vector to express the  $V_H$  and  $V_L$ -coding DNA homologs include the Shine-Dalgarno ribosome binding site, a leader sequence to direct the expressed protein to the periplasm as described by Mouva et al., *J. Biol. Chem.*, 255:27, 1980, and various restriction enzyme sites used to operatively link the  $V_H$  and  $V_L$  homologs to the expression vector. The  $V_H$  expression-vector sequence also contains a short nucleic acid sequence that codes for amino acids typically found in variable regions heavy chain ( $V_H$  Backbone). This  $V_H$  Backbone is just upstream and in the proper reading as the  $V_H$  DNA homologs that are operatively linked into the Xho I and Spe I. The  $V_L$  DNA homologs are operatively linked into the  $V_L$  sequence (6B) at the Nco I and Spe I restriction enzyme sites and thus the  $V_H$  Backbone region is deleted when the  $V_L$  DNA homologs are operatively linked into the  $V_L$  vector.

Figure 7 depicts the major features of the bacterial expression vector Lambda Zap II  $V_H$  ( $V_H$ -expression vector) are shown. The synthetic DNA sequence from Figure 6 is shown at the top along with the  $T_3$  polymerase promoter from Lambda Zap II vector. The orientation of the insert in Lambda Zap II vector is shown. The  $V_H$  DNA homologs are inserted into the Xho I and Spe I restriction enzyme sites. The  $V_H$  DNA are inserted into the Xho I and Spe I site and the read through transcription produces the

decapeptide epitope (tag) that is located just 3' of the cloning sites.

Figure 8 depicts the major features of the bacterial expression vector Lambda Zap II  $V_L$  ( $V_L$  expression vector) are shown. The synthetic sequence shown in Figure 6B is shown at the top along with the  $T_3$  polymerase promoter from Lambda Zap II vector. The orientation of the insert in Lambda Zap vector II is shown. The  $V_L$  DNA homologs are inserted into the phagemid that is produced by the in vivo excision protocol described by Short et al., Nucleic Acids Res., 16:7583-7600, 1988. The  $V_L$  DNA homologs are inserted into the Nco I and Spe I cloning sites of the Phagemid.

Figure 9 depicts a modified bacterial expression vector Lambda Zap II  $V_{LII}$ . This vector is constructed by inserting this synthetic DNA sequence,

TGAATTCTAACTAGTCGCCAAGGAGACAGTCATAATGAA  
TCGAACTTAAGATTTGATCAGCGGTTTCCTCTGTCAGTATTACTT  
ATACCTATTGCCTACGGCAGCCGCTGGATTGTTATTACTCGCTG  
TATGGATAACGGATGCCGTCGGCGACCTAACAATAATGAGCGAC  
20 CCCAACCAGCCATGGCCGAGCTCGTCAGTTCTAGAGTTAAGCGGCCG  
GGGTTGGTCGGTACCGGCTCGAGCAGTCAAGATCTCAATTCGCCGGCAGCT  
into Lambda Zap II vector that has been digested with the restriction enzymes Sac I and Xho I. This sequence contains the Shine-Dalgarno sequence (ribosome binding  
25 site), the leader sequence to direct the expressed protein to the periplasm and the appropriate nucleic acid sequence to allow the  $V_L$  DNA homologs to be operatively linked into the SacI and XbaI restriction enzyme sites provided by this vector.

30 Figure 10 depicts the sequence of the synthetic DNA segment inserted into Lambda Zap II vector to produce the lambda  $V_{LII}$ -expression vector. The various features and restriction endonuclease recognition sites are shown.

Figure 11 depicts the vectors for expressing  $V_H$  and  $V_L$  separately and in combination. The various essential components of these vectors are shown. The light chain vector or  $V_L$  expression vector can be combined with the  $V_H$

expression vector to produce a combinatorial vector containing both  $V_H$  and  $V_L$  operatively linked for expression to the same promoter.

Figure 12 depicts the labelled proteins immuno-precipitated from E. coli containing a  $V_H$  and a  $V_L$  DNA homolog are shown. In lane 1, the background proteins immunoprecipitated from E. coli that do not contain a  $V_H$  or  $V_L$  DNA homolog are shown. Lane 2 contains the  $V_H$  protein immunoprecipitated from E. coli containing only a  $V_H$  DNA homolog. In lanes 3 and 4, the commigration of a  $V_H$  protein a  $V_L$  protein immunoprecipitated from E. coli containing both a  $V_H$  and a  $V_L$  DNA homolog is shown. In lane 5 the presence of  $V_H$  protein and  $V_L$  protein expressed from the  $V_H$  and  $V_L$  DNA homologs is demonstrated by the two distinguishable protein species. Lane 5 contains the background proteins immunoprecipitated by anti-E. coli antibodies present in mouse ascites fluid.

Figure 13 depicts the transition state analogue (formula 1) which induces antibodies for hydrolyzing carboxamide substrate (formula 2). The compound of formula 1 containing a glutaryl spacer and a N-hydroxysuccinimide-linker appendage is the form used to couple the hapten (formula 1) to protein carriers KLH and BSA, while the compound of formula 3 is the inhibitor. The phosphoramidate functionality is a mimic of the stereoelectronic features of the transition state for hydrolysis of the amide bond.

Figure 14 illustrates the PCR amplification of Fd and kappa regions from the spleen mRNA of a mouse immunized with NPN. Amplification was performed as described in Example 17 using RNA cDNA hybrids obtained by the reverse transcription of the mRNA with primer specific for amplification of light chain sequences (Table 2) or heavy chain sequences (Table 1). Lanes F1-F8 represent the product of heavy chain amplification reactions with one of each of the eight 5' primers (primers 2-9, Table 1) and the unique 3' primer (primer 15, Table 2). Light chain (k) ampli-

cations with the 5' primers (primers 3-6, and 12, respectively, Table 2) are shown in lanes F9-F13. A band of 700 bps is seen in all lanes indicating the successful amplification of Fd and k regions.

5        Figure 15 depicts the screening of phage libraries for antigen binding is depicted according to Example 17C. Duplicate plaque lifts of Fab (filters A,B), heavy chain (filters E,F) and light chain (filters G,H) expression libraries were screened against  $^{125}\text{I}$ -labelled BSA conjugated  
10 with NPN at a density of approximately 30,000 plaques per plate. Filters C and D illustrate the duplicate secondary screening of a cored positive from a primary filter A (arrows) as discussed in the text.

Screening employed standard plaque lift methods. XL1  
15 Blue cells infected with phage were incubated on 150mm plates for 4 hours at 37°C, protein expression induced by overlay with nitrocellulose filters soaked in 10mM isopropyl thiogalactoside (IPTG) and the plates incubated at 25° for 8 hours. Duplicate filters were obtained during a  
20 second incubation employing the same conditions. Filters were then blocked in a solution of 1% BSA in PBS for 1 hour before incubation with rocking at 25° for 1 hour with a solution of  $^{125}\text{I}$ -labelled BSA conjugated to NPN ( $2 \times 10^6$  cpm  $\text{ml}^{-1}$ ; BSA concentration at 0.1 M; approximately 15 NPN  
25 per BSA molecule) in 1% BSA/PBS. Background was reduced by pre-centrifugation of stock radiolabelled BSA solution at 100,000 g for 15 minutes and pre-incubation of solutions with plaque lifts from plates containing bacteria infected with a phage having no insert. After labeling,  
30 filters were washed repeatedly with PBS/0.05% Tween 20 before development of autoradiographs overnight.

Figure 16 depicts the specificity of antigen binding as shown by competitive inhibition is illustrated according to Example 17C. Filter lifts from positive plaques  
35 were exposed to  $^{125}\text{I}$ -BSA-NPN in the presence of increasing concentrations of the inhibitor NPN.

In this study a number of phages correlated with NPN binding as in Figure 15 were spotted (about 100 particles per spot) directly onto a bacterial lawns. The plate was then overlaid with an IPTG-soaked filter and incubated for 19 hours at 25°. The filter were then blocked in 1% BSA in PBS prior to incubation in  $^{125}\text{I}$ -BSA-NPN as described previously in Figure 15 except with the inclusion of varying amounts of NPN in the labeling solution. Other conditions and procedures were as in Figure 15. The results for a phage of moderate affinity are shown in duplicate in the figure. Similar results were obtained for four other phages with some differences in the effective inhibitor concentration ranges.

Figure 17 depicts the characterization of an antigen binding protein is illustrated according to Example 17D. The concentrated partially purified bacterial supernate of an NPN-binding clone was separated by gel filtration and aliquots from each fraction applied to microtitre plates coated with BSA-NPN. Addition of either anti-decapeptide (---) or anti-kappa chain antibodies conjugated with alkaline phosphatase was followed by color development. The arrow indicates the position of elution of a known Fab fragment. The results show that antigen binding is a property of 50 kD protein containing both heavy and light chains.

Single plaques of two-NPN-positive clones (Figure 15) were picked and the plasmid containing the heavy and light chain inserts excised. 500 ml cultures in L-broth were inoculated with 3 ml of a saturated culture containing the excised plasmids and incubated for 4 hours at 37°C. Proteins synthesis was induced by the addition of IPTG to a final concentration of 1mM and the cultures incubated for 10 hours at 25°C. 200 ml of cells supernate were concentrated to 2 ml and applied to a TSK-G4000 column. 50  $\mu\text{l}$  aliquots from the eluted fractions were assayed by ELISA.



For ELISA analysis, microtitre plates were coated with BSA-NPN at 1 ug/ml, 50  $\mu$ l samples mixed with 50  $\mu$ l PBS-Tween 20 (0.05%)-BSA (0.1%) added and the plates incubated for 2 hours at 25°. After washing with PBS-Tween 20-BSA, 50  $\mu$ l of appropriate concentrations of a rabbit anti-decapeptide antibody (20) and a goat anti-mouse kappa light chain (Southern Biotech) antibody conjugated with alkaline phosphatase were added and incubated for 2 hours at 25°. After further washing, 50  $\mu$ l of p-nitrophenyl phosphate (1mg/ml in 0.1M Tris pH 9.5 containing 50 mM MgCl<sub>2</sub>) were added and the plates incubated for 15-30 minutes before reading the OD at 405nm.

Figure 18A depicts the major features of the bacterial expression vector HCFLP containing a V<sub>H</sub> DNA homolog and a flp recombination site.

Figure 18B depicts the major features of the bacterial expression vector LCFLP containing a V<sub>L</sub> DNA homolog and a flp recombination site properly oriented for recombination with the HCFLP vector.

Figure 19 depicts a diagrammatic sketch of bacterial coinfection with HCFLP and LCFLP vectors for the production of recombinant expression vectors containing V<sub>L</sub> and V<sub>H</sub> DNA homologs.

Figure 20 depicts an outline showing arm selection for heavy and light chain recombinant vector products using flp recombinase in conjunction with selection based on the inclusion of genes having amber mutations.

Figure 21 shows an outline of a method of phenotype creation using the fusion PCR process described herein.

Figure 22 illustrates human fusion PCR inside primers. The heavy chain C<sub>H</sub>1' inside primer sequence is written 3' to 5' and the light chain V<sub>L</sub> inside primer sequence is written 5' to 3'. Note that it is not the primer strands that cross-prime to create the fusion molecule, but the complementary PCR product strands. Boxed nucleotides represent regions where the C<sub>H</sub>1' primer hybridizes to the 3' end of C<sub>H</sub>1 on human IgG heavy chain

mRNA or where the  $V_L$  primer hybridizes to the 5' end of  $V_L$  framework-1 on human kappa light chain cDNA. Underlined sequences indicate the two stop condons. The italicized amino acid and nucleotides indicate changes in sequence from the original pelB leader sequence. The mouse fusion-PCR internal primers overlap in a similar manner.

Figure 23 illustrates an ethidium bromide stained agarose gel. After PCR amplification from human cloned DNA of heavy chain alone (HC), light chain alone (LC), and the heavy/light dicistronic DNA molecule (H/L), DNA samples were electrophoresed. The expected sizes of the HC, LC, and H/L products visualized on the gel were approximately 730, 690, and 1,390 base pairs, respectively.

Figures 24A and 24B illustrate the major features of the bacterial expression vector Lambda ZAP II Modified  $V_H$  (Modified ImmunoZAP H) ( $V_H$ -expression vector) (IZ H). The amino acids encoded by the synthetic DNA sequence from Figure 24A is shown along with the  $T_3$  polymerase promoter from Lambda ZAP II. The orientation of the insert in Lambda ZAP II is as presented. The insert was modified by the elimination of the Sac I site between the  $T_3$  polymerase and Not I site and by the change of amino acids at the 5' end of the heavy chain from QVKL to QVQL (alysine residue was changed to a glutamine residue). The  $V_H$  and  $V_L$  DNA homologs were inserted into the Xho I and Xba I cloning sites of the phagemid as described in Figure 26 and shown in Figure 24B. The modifications were made to create a fusion-PCR library from hybridoma RNA, to overcome decreased efficiency of secretion of positively charged amino acids in the amino terminus of the protein. Inouye et al., Proc. Natl. Acad. Sci., USA, 85:7685-7689 (1988), and to make the  $V_L$  Sac I cloning site a unique restriction site.

Figures 25A and 25B illustrate the sequences of the synthetic DNAs inserted into Lambda ZAP to produce Lambda Zap II  $V_H$  (ImmunoZAP H) (25A) and Lambda Zap  $V_L$  (ImmunoZAP L) (25B) expression vectors. The various features

required for these vectors to express the  $V_H$  and  $V_L$ -coding DNA homologs include the Shine-Dalgarno ribosome binding site, a leader sequence to direct the expressed protein to the periplasm as described by Mouva et al., J. Biol. Chem., 255:27, 1980, and various restriction enzyme sites used to operatively link the  $V_H$  and  $V_L$  homologs to the expression vector. The  $V_H$  expression-vector sequence also contains a short nucleic acid sequence that codes for amino acids typically found in variable regions of the heavy chain ( $V_H$  Backbone). This  $V_H$  Backbone is just upstream and in the proper reading frame as the  $V_H$  DNA homologs that are operatively linked into the Xho I and Spe I restriction sites. The  $V_L$  DNA homologs are operatively linked into the  $V_L$  sequence (25B) at the Sac I and Xba I restriction enzyme sites.

Figure 26 illustrates the major features of the bacterial expression vector Lambda Zap II  $V_H$  (ImmunoZAP H) ( $V_H$ - expression vector). The amino acids encoded by the synthetic DNA sequence from Figure 25A is shown at the top along with the  $T_3$  polymerase promoter from Lambda Zap II. The orientation of the insert in Lambda Zap II is as presented. The  $V_H$  DNA homologs were inserted into the phagemid that is produced by the in vivo excision protocol described by Short et al., Nucleic Acids Res., 16:7583-7600, 1988. The  $V_H$  DNA homologs were inserted into the Xho I and Spe I restriction enzyme sites. The read through transcription produces the decapeptide epitope (tag) that is located just 3' of the cloning sites.

Figure 27 illustrates the major features of the bacterial expression vector Lambda Zap II  $V_L$  (ImmunoZAP L) ( $V_L$  expression vector). The amino acids encoded by the synthetic DNA sequence shown in Figure 25B is shown at the top along with the  $T_3$  polymerase promoter from Lambda Zap II. The orientation of the insert in Lambda Zap II is as presented. The  $V_L$  DNA homologs are inserted into the Sac I and Xba I cloning sites of the phagemid as described in Figure 26.

Figure 28 illustrated an autoradiogram showing signals obtained from human phage clones. Approximately 100 lambda phage were spotted onto E. coli lawns, creating plaques that were overlaid with nitrocellulose filters previously soaked in 10 mM isopropylbeta-D-thiogalactopyranoside (IPTG) to induce Fab expression. Following overnight incubation, the filters were reacted with <sup>125</sup>I-tetanus toxoid probe. After washing, the filters were exposed to X-ray film. The column on the right represents the parental clones that were selected from a combinatorial library. Mullinax et al., Proc. Natl. Acad. Sci., USA, 87:8095-8099 (1990). The column on the left represents clones that were generated by amplifying the combinatorial lambda clone DNA with the V<sub>H</sub> and C<sub>L</sub>' outside primers, C<sub>H</sub>1' and V<sub>L</sub> inside primers, followed by recloning in the modified ImmunoZAP H vector. Clone 7G1 is a negative control which expresses a Fab that does not react with tetanus toxoid. Clones 10C1 and 6C12 both produce Fabs that react with tetanus toxoid. IZ H is the modified heavy chain ImmunoZAP H vector without an insert.

### Detailed Description of the Invention

#### A. Definitions

As used herein, the following terms have the following meanings unless expressly stated to the contrary:

Nucleotide: a monomeric unit of DNA or RNA consisting of a sugar moiety (pentose), a phosphate, and a nitrogenous heterocyclic base. The base is linked to the sugar moiety via the glycosidic carbon (1' carbon of the pentose) and that combination of base and sugar is a nucleoside. When the nucleoside contains a phosphate group bonded to the 3' or 5' position of the pentose it is referred to as a nucleotide.

Base Pair (bp): a pairing (by hydrogen bonding) of adenine (A) with thymine (T), or of cytosine (C) with guanine (G) in a double stranded DNA molecule. In RNA, uracil (U) is substituted for thymine.

**SUBSTITUTE SHEET**

Nucleic Acid: a polymer of nucleotides, either single or double stranded.

Gene: a nucleic acid whose nucleotide sequence codes for an RNA or polypeptide. A gene can be either RNA or  
5 DNA.

Complementary Bases: nucleotides that normally pair up when DNA or RNA adopts a double stranded configuration.

Complementary Nucleotide Sequence: a sequence of nucleotides in a single-stranded molecule of DNA or RNA  
10 that is sufficiently complementary to that on another single strand to specifically hybridize to it with consequent hydrogen bonding.

Conserved: a nucleotide sequence is conserved with respect to a preselected (reference) sequence if it non-  
15 randomly hybridizes to an exact complement of the preselected sequence.

Hybridization: the pairing of substantially complementary nucleotide sequences (strands of nucleic acid) to form a duplex or heteroduplex by the establish-  
20 ment of hydrogen bonds between complementary base pairs. It is a specific, i.e. non-random, interaction between two complementary polynucleotides that can be competitively inhibited.

Nucleotide Analog: a purine or pyrimidine nucleotide  
25 that differs structurally from A, T, G, C, or U, but is sufficiently similar to substitute for the normal nucleotide in a nucleic acid molecule.

DNA Homolog: is a nucleic acid having a preselected conserved nucleotide sequence and a sequence coding for a  
30 receptor capable of binding a preselected ligand.

Receptor: A receptor is a molecule, such as a protein, glycoprotein and the like, that can specifically (non-randomly) bind to another molecule.

Antibody: The term antibody in its various grammatical forms is used herein to refer to immunoglobulin  
35 molecules and immunologically active portions of immunoglobulin molecules, i.e., molecules that contain an

antibody combining site or paratope. Exemplary antibody molecules are intact immunoglobulin molecules, substantially intact immunoglobulin molecules and portions of an immunoglobulin molecule, including those portions known in the art as Fab, Fab', F(ab')<sub>2</sub> and F(v).

Antibody Combining Site: An antibody combining site is that structural portion of an antibody molecule comprised of a heavy and light chain variable and hypervariable regions that specifically binds (immunoreacts with) an antigen. The term immunoreact in its various forms means specific binding between an antigenic determinant-containing molecule and a molecule containing an antibody combining site such as a whole antibody molecule or a portion thereof.

Monoclonal Antibody: The phrase monoclonal antibody in its various grammatical forms refers to a population of antibody molecules that contains only one species of antibody combining site capable of immunoreacting with a particular antigen. A monoclonal antibody thus typically displays a single binding affinity for any antigen with which it immunoreacts. A monoclonal antibody may therefore contain an antibody molecule having a plurality of antibody combining sites, each immunospecific for a different antigen, e.g., a bispecific monoclonal antibody.

Upstream: In the direction opposite to the direction of DNA transcription, and therefore going from 5' to 3' on the non-coding strand, or 3' to 5' on the mRNA.

Downstream: Further along a DNA sequence in the direction of sequence transcription or read out, that is traveling in a 3'- to 5'-direction along the non-coding strand of the DNA or 5'- to 3'-direction along the RNA transcript.

Cistron: Sequence of nucleotides in a DNA molecule coding for an amino acid residue sequence.

Stop Codon: Any of three codons that do not code for an amino acid, but instead cause termination of protein

synthesis. They are UAG, UAA and UGA. Also referred to as a nonsense or termination codon.

Leader Polypeptide: A short length of amino acid sequence at the amino end of a protein, which carries or  
5 directs the protein through the inner membrane and so ensures its eventual secretion into the periplasmic space and perhaps beyond. The leader sequence peptide is commonly removed before the protein becomes active.

Reading Frame: Particular sequence of contiguous  
10 nucleotide triplets (codons) employed in translation. The reading frame depends on the location of the translation initiation codon.

Inside Primer: An inside primer is a polynucleotide that has a priming region located at the 3' terminus of  
15 the primer which typically consists of 15 to 30 nucleotide bases. The 3' terminal-priming portion is capable of acting as a primer to catalyze nucleic acid synthesis. The 5'-terminal priming portion comprises a non-priming portion.

20 Outside Primer: An outside primer comprises a 3'-terminal priming portion and a portion that may define an endonuclease restriction site which is typically located in a 5'-terminal non-priming portion of the outside primer.

25 Fusion Polynucleotide Amplification: refers to in vitro techniques of generating a multiple complementary copies of a nucleic acid template which comprises nucleotide sequences which have been randomly combined to give a combined nucleic sequence. These techniques typically  
30 employ complementary primers which hybridize to the template and are extended in a primer extension reaction. The polymerase chain reaction (PCR) techniques described herein comprise a preferred method of nucleotide sequence amplifications. Generation and amplification of a  
35 combined nucleotide sequence using fusion PCR is further described herein.

Vector: As used herein, the term "vector" refers to a nucleic acid molecule capable to transporting between different genetic environments another nucleic acid to which it has been operatively linked. One type of preferred vector is an episome, i.e., a nucleic acid molecule capable of extra-chromosomal replication. Other suitable vectors include plasmid and cosmid vectors and phage, especially bacteriophage such as lambda. Preferred vectors are those capable of autonomous replication and/or expression of nucleic acids to which they are linked. Vectors capable of directing the expression of genes to which they are operatively linked are referred to herein as "expression vectors".

B. Methods

Until this invention, genetic engineers typically dealt with the expression of a single gene or family (or population) of genes, one at a time. The expression of a family of genes in a vector is generally referred to as a "gene library." Each member of the library will normally contain a different gene or DNA sequence. However, the vector portion of such a vector-gene fusion is typically identical from member to member. (Maniatis et al., supra). Individual members within the library may often be, and typically are, amplified before screening to identify and isolate a desired member. Amplification occurs so that each library member grows as a bacterial colony (for plasmid libraries) or phage plaques (for bacteriophage libraries, such as lambda). These amplified members are usually referred to as "clones," since each colony or plaque is made up of many identical host cells or phage particles.

The search for a particular clone containing a single gene or DNA sequence of interest can be accomplished in many different ways. The clone may be identified because its vector-gene specifically hybridizes with a nucleic acid probe. It may also be identified by expression of an



RNA species that can be identified, for example by nucleic acid hybridization. The RNA species may, furthermore, be translated into a protein, typically by the host cell, that may be identified, for example, by reactivity with an antibody probe. Alternatively, the protein may be recognized because it binds a substrate, or catalyzes a reaction, or allows the host cell to survive under selective conditions, and so on.

Described herein are libraries in which two or more families (or populations) of genes are expressed in a vector or a host cell in such a way that the gene combinations are randomly represented and subsequently detected on the basis of some property or characteristic in the event that a particular combination of one member from a first gene family and one gene from a one or more other gene families are combined in a vector host cell. For example, in the general case if there are "i" members of the gene family "A" and "j" members of the gene family "B", there will be  $(i) \times (j)$  combinations of selected gene members A and B in the randomly created vector-gene population. If there are three gene families, A, B, and C, and a vector is made containing one member from each of the three gene families, the total number of combinations of genes will be the product of the number of A genes times the number of B genes times the number of C genes. Thus, methods are provided wherein at least two genes may randomly be combined, preferably on the same vector molecule, having been identified within a population of vectors containing other combinations of different genes from the same two or more gene families. This approach may be broadly accomplished by means other than recombination, for example, the use of a vector having at least two independent insertion sites for two foreign genes or inserting in a vector a nucleotide sequence comprising nucleotide sequences from each gene family. The recombination of at least two separate library populations to make a combinatorial population, for example, using a

common restriction site or site-directed recombination systems, is also contemplated.

Thus, in addition to the above-described methods, the invention also provides for vectors having characteristics and sequences useful for the preparation of combinatorial vectors encoding random DNA sequences from two or more gene families. Such vectors include plasmids and phage containing common restriction sites or sequences enabling the in vivo recombination of said DNA sequences from said gene families.

The flp site-specific recombination of S. cerevisiae has been described in Cox, Chapter 13 in "Genetic Recombination," eds. R. Kucherlapati and G. Smith (American Society for Microbiology 1988). Within a sixty-five bp region identified as the recombination site and designated FRT (flp recombination target), there are several prominent structural features. The most important are a set of three bp repeats. The second and third repeats are separated by one bp and are in the same orientation. The first repeat is inverted with respect to the other two and is separated from the second repeat by an eight bp spacer. The first repeat also has a one bp mismatch relative to the first two. Deletion analysis has demonstrated that the third repeat is unnecessary for recombination in vitro, although it may have a slight effect on the reaction in vivo. Additional deletions indicate that most, but not all, of the first and second repeats (those flanking the spacer) are required. While deletion of three bp from the distal ends of one or both of these repeats has no detectable effect on the reaction, further deletion leads to a gradual reduction in site function, with complete loss of site function occurring (in vitro) with deletions of eight bp or more from either end. The minimal site required for a full function in vitro is therefore relatively small (approximately 28 bp including the spacer and the proximal 10 bp of each flanking repeat). Accordingly, it will be seen that the full,

intermediate, or minimal FRT sequences can be utilized to accomplish flp-mediated site-specific recombination.

The lambda phage attachment site is responsible for integration of lambda into the host chromosome. It also  
5 acts as a hot spot of recombination and lytic crosses between wild lambda chromosomes. As in lambda, in P1 phage a site-specific cross over site, loxP acts as a hot spot of recombination. This site is recognized by the P1 cre protein, a known site-specific protein. The site-  
10 specific recombination system is responsible for the rare integration of P1 into the host chromosome. The cre-lox system of bacteriophage P1 is also useful for the site-specific recombination contemplated by the invention described and claimed herein.

15 A transposon can jump from one vector to another vector or from a vector to a bacterial chromosome. Different transposons having different inverted repeat sequences and carrying, for example, different drug-resistance genes, can be used to carry out the desired  
20 random combination of genes as described herein either in vivo or in vitro. The transposon may, but need not, also contain a sequence encoding the transposase enzyme which catalyzes the "hop." Various suitable transposon systems have been described in the literature. (See, Mobile DNA,  
25 Douglas E. Berg and Martha M. Howe, eds., American Society for Microbiology, Washington, D.C., 1989). One suitable transposon system is the gamma-delta transposon system which has been isolated from E. Coli.

Thus, in addition to restriction digestion and  
30 ligation, use of flp type recombination systems, and homologous recombination, a transposon system can also be used to integrate a light (or heavy) antibody chain clone into a heavy (or light) antibody chain clone. For example, this can be accomplished by flanking the light  
35 chain expression and cloning region with transposon terminal sequences. A library constructed in this light chain vector could be used to co-infect bacteria with

clones from the heavy chain library. The light chain inserts between the terminal sequences would hop from the light chain lambda phage vector into other DNA sequences in the presence of transposase activity. Selection for hopping into the heavy chain clone can be accomplished by placing a selectable marker within the light chain, positioned between the transposon hopping sequences. Subsequently, phage recovered from the co-infected culture is plated with a strain enabling selection for the heavy chain vector and for the light chain marker gene. Because this second plating is performed under conditions of a high cell to phage ratio, only one lambda phage will typically be introduced into each cell. The lambda phage should grow only if the phage contains genes from both the heavy and light chain clones; most efficiently resulting from the transposon hop. If the hop occurs in the essential genes of the heavy chain clone, the phage will not grow. Only phage containing the transposon in the proper position within the heavy chain will grow. A collection of these clones comprises a library of combinatorial heavy and light chain antibody clones.

According to one aspect of the present invention, fusion PCR is used to generate two PCR-amplified DNA fragments, each of which have one of their ends modified by directed mispriming so that those ends share regions of complementarity, i.e., cohesive termini. When the two fragments are mixed, denatured and reannealed in a PCR cycle, the cohesive termini on two strands hybridize to form an "overlapping" DNA duplex that is internally primed. The subsequent PCR cycle primer-extends the non-overlapping regions to form a hybrid DNA molecule that is dicistronic. See Figure 21.

PCR amplification methods are described in detail in U.S. Patent Nos. 4,863,192, 4,683,202, 4,800,159, and 4,965,188, and at least in several texts including "PCR Technology: Principles and Applications for DNA Amplification", H. Erlich, ed., Stockton Press, New York

(1989); and "PCR Protocols" A Guide to Methods and Applications", Innis et al., eds., Academic Press, San Diego, California (1990).

Thus, in one aspect of the present invention, fusion  
5 PCR is used to produce a library of dicistronic DNA molecules containing upstream and downstream cistrons wherein first and second PCR amplification products are produced using respective first and second PCR primer pairs. The first PCR primer pair comprises a first polypeptide  
10 outside primer and a first polypeptide inside primer. Similarly, the second PCR primer pair comprises a second polypeptide outside primer and a second polypeptide inside primer. The first and second polypeptide inside primers contain complementary 5'-terminal sequences that allow  
15 their DNA complements to hybridize and form an internally-primed duplex having 3'-overhanging termini. The internally-primed duplex is then subjected to primer extension reaction conditions to produce a double stranded, dicistronic DNA having substantially blunt or  
20 blunt ends. The dicistronic DNA is then PCR amplified using the outside primers as a PCR primer pair.

The dicistronic DNA molecule comprises two amino acid residue-coding sequences on the same strand separated by at least one stop codon and at least one signal sequence  
25 necessary for translation of the downstream cistron, such as a translation initiation codon, ribosome binding site, and the like. Thus, the upstream and downstream cistrons of the dicistronic DNA molecule are operatively linked by a cistronic bridge. The cistronic bridge comprises the  
30 genetic elements necessary to terminate translation of the upstream cistron and initiate translation of the downstream cistron. For instance, the coding strand of the bridge codes for one or more stop codons, preferably two, in the same translational reading frame as the upstream  
35 cistron. The cistronic bridge coding strand preferably also encodes a ribosome binding site for the downstream cistron located downstream from the upstream cistron's

stop codon(s). Typically, the coding strand of the cistronic bridge will also encode a leader polypeptide segment in the same translational reading frame as the downstream cistron. When present, the nucleotide base sequence encoding the leader usually begins with an initiation codon located within an operative distance, i.e. is operatively linked, to the ribosome binding site.

The following discussion illustrates the use of fusion PCR to isolate a pair of  $V_H$  and  $V_L$  genes from the immunoglobulin gene repertoire. This discussion is not to be taken as limiting, but rather as illustrating an application of creating a novel phenotype by combining one member from each of two or more families of genes. The illustrated method can be used with other families of conserved genes which each for one unit of a dimeric receptor, whether obtained directly from a natural source, such naive or in vivo immunized cells, or from cells or one or more genes that have been treated or mutagenized in vitro. Generally, the method, combines the following elements:

1. Producing  $V_H$  and  $V_L$  gene repertoires.
2. Preparing sets of outside and inside polynucleotide primers for cloning polynucleotide segments containing immunoglobulin  $V_H$  and  $V_L$  region genes.
- 25 3. Preparing a library containing a plurality of different dicistronic DNA molecules, each containing a  $V_H$  and a  $V_L$  gene from the respective repertoires.
4. Expressing the dicistronic DNA molecules in suitable host cells.
- 30 5. Screening the polypeptides expressed by the dicistronic DNA molecules for the preselected activity, and segregating a dicistronic DNA molecules for the preselected activity, and segregating a dicistronic DNA molecule identified by the screening process.

35 The present invention also provides a novel method for screening variants of a parental clone or clones. If the parental clone or clones contain two

nucleotide sequences that, when expressed together, create a phenotype, then such nucleotide sequences can be altered to create populations of variants of such nucleotide sequences. If the two variant populations are coexpressed  
5 in a random fashion (that is with no correlation between the specific alterations made in the two different nucleotide sequences), then a combinatorial collection of such nucleotide sequence variants has been created. Such combinatorial collections may be screened for the presence  
10 of phenotypes that are unlike the parental clone or clones. Generally, the method combines the following elements:

1. Replicating a clone containing a nucleotide sequence under conditions that allow mutations to occur.
- 15 2. Replicating a second clone containing a second nucleotide sequence under conditions that allow mutations to occur.
3. Randomly combining and co-expressing the two mutated populations of nucleotide sequences.
- 20 4. Screening clones containing combinations of mutated nucleotide sequences for phenotypes that were not present in either parent clone.

Alternatively, the methods combine the following elements:

- 25 1. Replicating at least portions of two nucleotide sequences contained within a single clone under conditions that allow mutations to occur in either nucleotide sequence.
2. Allowing recombination events between the two  
30 nucleotide sequence populations to reassociate mutant nucleotide sequences to form new pairs of the two sequences that were not paired in the original mutated, replicated population.
3. Screening clones containing combinations of  
35 nucleotide sequences for phenotypes that were not present in the parent clone or in the mutant replicas of the parent clone.

For example, assume a parent clone containing two nucleotide sequences A and B is replicated under mutating conditions such that variant clones are formed:

Parent: A/B

5 Variant 1: A1/B

Variant 2: A/B1

Variant 3: A2/B1

Variant 4: A/B2

Variant 5: A3/B

10 However, within this mutated population, the combinations A1/B2, A2/B, A2/B2, A3/B1, and A3/B2, do not occur. If the mutant population (including some non-mutated parent clones) is allowed to recombine sequences A and B and their variants, then combinations such as A1/B2, A2/B etc.  
15 can be created. Such new combinations may express a desired phenotype that was not present in the parental or the variant population.

In one aspect, the present invention is related to methods for tapping the immunological repertoire by  
20 isolating from  $V_H$ -coding and  $V_L$ -coding gene repertoires genes coding for a heterodimeric antibody receptor capable of binding a preselected ligand. Generally, the method combines the following elements:

1. Isolating nucleic acids containing a substantial  
25 portion of the immunological repertoire.

2. Preparing polynucleotide primers for cloning polynucleotide segments containing immunoglobulin  $V_H$  and  $V_L$  region genes.

3. Preparing a gene library containing a plurality  
30 of different  $V_H$  and  $V_L$  genes from the repertoire.

4. Expressing the  $V_H$  and  $V_L$  polypeptides in a suitable host, including prokaryotic and eukaryotic hosts, on the same expression vector.

5. Screening the expressed polypeptides for the  
35 preselected activity, and segregating a  $V_H$ - and  $V_L$ -coding gene combination identified by the screening process.



In one aspect, the expressed phenotype produced by the methods by the present invention comprises a multimeric polypeptide product (i.e. a heterodimer, etc.) which assumes a conformation having a binding site specific for, as evidenced by its ability to be competitively inhibited, a preselected or predetermined ligand such as an antigen, enzymatic substrate and the like. In one embodiment, the multimeric polypeptide is an antibody that forms an antigen binding site which specifically binds to a preselected antigen to form an immunoreaction product (complex) having a sufficiently strong binding between the antigen and the binding site for the immunoreaction product to be isolated. The antibody typically has an affinity or avidity is generally greater than  $10^5\text{-M}^{-1}$ .

In another embodiment, a multimeric polypeptide produced according to the present invention is capable of binding a substrate and catalyzes the formation of a product from the substrate. While the topology of the ligand binding site of a catalyzing multimeric polypeptide is probably more important for its preselected activity than its affinity (association constant or pKa) for the substrate, the useful catalytic multimeric polypeptides typically have an association constant for the preselected substrate generally greater than  $10^3\text{ M}^{-1}$ , more usually greater than  $10^5\text{ M}^{-1}$  or  $10^6\text{ M}^{-1}$  and preferably greater than  $10^7\text{ M}^{-1}$ .

Preferably the multimeric polypeptide produced according to the present invention is heterodimeric and is therefore normally comprised of two different polypeptide chains, which together assume a conformation having a binding affinity, or association constant for the preselected ligand that is different, preferably higher, than the affinity or association constant of either of the polypeptides alone, i.e., as monomers. In a particularly preferred aspect, one or both of the different polypeptide chains is derived from the variable region of the light and heavy chains of an immunoglobulin. Typically, poly-

peptides comprising the light ( $V_L$ ) and heavy ( $V_H$ ) variable regions are employed together for binding the preselected ligand.

A  $V_H$  or  $V_L$  produced by the methods of the subject invention can be active in monomeric as well as multimeric forms, either homomeric or heteromeric, preferably heterodimeric. A  $V_H$  and  $V_L$  ligand binding polypeptide produced by the present invention can be advantageously combined in a heterodimer (antibody molecule) to modulate the activity of either or to produce an activity unique to the heterodimer. The individual ligand binding polypeptides will be referred to as  $V_H$  and  $V_L$  and the heterodimer will be referred to as an antibody molecule.

However, it should be understood that a  $V_H$  binding polypeptide may contain in addition to the  $V_H$ , substantially all or a portion of the heavy chain constant region. A  $V_L$  binding polypeptide may contain, in addition to the  $V_L$ , substantially all or a portion of the light chain constant region. A heterodimer comprised of a  $V_H$  binding polypeptide containing a portion of the heavy chain constant region and a  $V_L$  binding containing substantially all of the light chain constant region is termed a Fab fragment. The production of a Fab can be advantageous in some situations because the additional constant region sequences contained in a Fab as compared to a  $F_v$  could stabilize the  $V_H$  and  $V_L$  interaction. Such stabilization could cause the Fab to have higher affinity for antigen. In addition the Fab is more commonly used in the art and thus there are more commercial antibodies available to specifically recognize a Fab.

The individual  $V_H$  and  $V_L$  polypeptides may be produced in lengths equal or substantially equal to their naturally occurring lengths. However, the individual  $V_H$  and  $V_L$  polypeptides will generally have fewer than 125 amino acid residues, more usually fewer than about 120 amino acid residues, while normally having greater than 60 amino acid residues, usually greater than about 95 amino acid

residues, more usually greater than about 100 amino acid residues. Preferably, the  $V_H$  will be from about 110 to about 125 amino acid residues in length while  $V_L$  will be from about 95 to about 115 amino acid residues in length.

5       The amino acid residue sequences of the polypeptides will vary widely, depending upon the particular idio-  
type involved. Usually, there will be at least two cysteines separated by from about 60 to 75 amino acid residues and joined by a disulfide bond. The polypeptides produced by  
10   the subject invention will normally be substantial copies of idiotypes of the variable regions of the heavy and/or light chains of immunoglobulins, but in some situations a polypeptide may contain random mutations in amino acid residue sequences in order to advantageously improve the  
15   desired activity.

      In some situations, it is desirable to provide for covalent cross linking of the  $V_H$  and  $V_L$  polypeptides, which can be accomplished by providing cysteine residues at the carboxyl termini. The polypeptide will normally be pre-  
20   pared free of the immunoglobulin constant regions, however a small portion of the J region may be included as a result of the advantageous selection of DNA synthesis primers. The D region will normally be included in the transcript of the  $V_H$ .

25       In other situations, it is desirable to provide a peptide linker to connect the  $V_L$  and the  $V_H$  to form a single-chain antigen-binding protein comprised of a  $V_H$  and a  $V_L$ . This single-chain antigen-binding protein would be synthesized as a single protein chain. Such a single-  
30   chain antigen binding proteins have been described by Bird et al., Science, 242:423-426 (1988). The design of suitable peptide linker regions is described in U.S. Patent No. 4,704,692 by Robert Landner.

      Such a peptide linker may be designed as part of the  
35   nucleic acid sequences contained in the expression vector. The nucleic acid sequences coding for the peptide linker would be between the  $V_H$  and  $V_L$  DNA homologs and the

restriction endonuclease sites used to operatively link the  $V_H$  and  $V_L$  DNA homologs to the expression vector.

Such a peptide linker also may be coded for nucleic acid sequences that are part of the polynucleotide primers used to prepare the various gene libraries. The nucleic acid sequence coding for the peptide linker can be made up of nucleic acids attached to one of the primers or the nucleic acid sequence coding for the peptide linker may be derived from nucleic acid sequences that are attached to several polynucleotide primers used to create the gene libraries.

Typically the C terminus region of the  $V_H$  and  $V_L$  polypeptides will have a greater variety of the sequences than the N terminus and, based on the present strategy, can be further modified to permit a variation of the normally occurring  $V_H$  and  $V_L$  chains. A synthetic polynucleotide and be employed by vary one or more amino in an hypervariable region.

#### 1. Isolation Of A Gene Repertoire

According to one aspect of the present invention, a gene repertoire useful in the methods the present invention contains at least  $10^3$ , preferably at least  $10^4$ , more preferably at least  $10^5$ , and most preferably at least  $10^7$  different conserved genes. Methods for evaluating the diversity of a repertoire of conserved genes are well known to one skilled in the art.

Various well known methods can be employed to produce a useful gene repertoire. For example, to prepare a composition of nucleic acids containing a substantial portion of the immunological gene repertoire, a source of genes coding for the  $V_H$  and/or  $V_L$  polypeptides is required. Preferably the source will be heterogeneous population of antibody producing cells, *i.e.*, B lymphocytes (B cells), preferably rearranged B cells such as those found in the circulation or spleen of a vertebrate. (Rearranged B cells are those in which immunoglobulin gene transloca-

tion, i.e., rearrangement, has occurred as evidenced by the presence in the cell of mRNA with the immunoglobulin gene V, D and J region transcripts adjacently located thereon.)

5        In some cases, it is desirable to bias the repertoire for a preselected activity, such as by using as a source of nucleic acid cells (source cells) from vertebrates in any one of various stages of age, health and immune response. For example, repeated immunization of a healthy  
10 animal prior to collecting rearranged B cells results in obtaining a repertoire enriched for genetic material producing a ligand binding polypeptide of high affinity. See, e.g. Mullinax et al., Proc. Nat. Acad. Sci. (USA) 87:8095-8099 (1990). Conversely, collecting rearranged B  
15 cells from a healthy animal whose immune system had not been recently challenged results in producing a repertoire that is not biased towards the production of high affinity  $V_H$  and/or  $V_L$  polypeptides.

It should be noted the greater the genetic hetero-  
20 geneity of the population of cells for which the nucleic acids are obtained, the greater the diversity of the immunological repertoire that will be made available for screening according to the method of the present invention. Thus, cells from different individuals of different  
25 strains, races or species can be advantageously combined to increase the heterogeneity (diversity) of the repertoire.

Thus, in one preferred embodiment, the source cells are obtained from a vertebrate, preferably a mammal, which  
30 has been immunized or partially immunized with an antigenic ligand (antigen) against which activity is sought, i.e., a preselected antigen. The immunization can be carried out conventionally. Antibody titer in the animal can be monitored to determine the stage of immunization  
35 desired, which stage corresponds to the amount of enrichment or biasing of the repertoire desired. Partially immunized animals typically receive only one immunization

and cells are collected therefrom shortly after a response is detected. Fully immunized animals display a peak titer, which is achieved with one or more repeated injections of the antigen into the host mammal, normally at 2 to 3 week intervals. Usually three to five days after the last challenge, the spleen is removed and the genetic repertoire of the spleenocytes, about 90% of which are rearranged B cells, is isolated using standard procedures. See, Current Protocols in Molecular Biology, Ausubel et al., eds., John Wiley & Sons, NY.

Nucleic acids coding for  $V_H$  and  $V_L$  polypeptides can be derived from cells producing IgA, IgD, IgE, IgG or IgM, most preferably from IgM and IgG, producing cells.

Methods for preparing fragments of genomic DNA from which immunoglobulin variable region genes can be cloned as a diverse population are well known in the art. See for example Herrmann et al., Methods In Enzymol., 152:180-183, (1987); Frischauf, Methods In Enzymol., 152:180-190 (1987); Frischauf, Methods In Enzymol., 152:190-199 (1987); and DiLella et al., Methods In Enzymol., 152:199-212 (1987). (The teachings of the references cited herein are hereby incorporated by reference.)

The desired gene repertoire can be isolated from either genomic material containing the gene expressing the variable region or the messenger RNA (mRNA) which represents a transcript of the variable region. The difficulty in using the genomic DNA from other than non-rearranged B lymphocytes is in juxtaposing the sequences coding for the variable region, where the sequences are separated by intervening regions. The DNA fragment(s) containing the proper variable regions must be isolated, the intervening regions excised, and the variable regions then spliced in the proper order and in the proper orientation. For the most part, this will be difficult, so that the alternative technique employing rearranged B cells will be the method of choice because the V, D and J immunoglobulin gene

regions have translocated to become adjacent, so that the sequence is continuous for the variable regions.

Where mRNA is utilized the cells will be lysed under RNase inhibiting conditions. In one embodiment, the first  
5 step is to isolate the total cellular mRNA by hybridization to an oligo-dT cellulose column. The presence of mRNAs coding for the heavy and/or light chain polypeptides can then be assayed by hybridization with DNA single strands of the appropriate genes. Conveniently, the  
10 sequences coding for the constant portion of the  $V_H$  and  $V_L$  can be used as polynucleotide probes, which sequences can be obtained from available sources. See for example, Early and Hood, Genetic Engineering, Setlow and Hollaender, eds., Vol. 3, Plenum Publishing Corporation,  
15 New York, (1981), pages 157-188; and Kabat et al., Sequences of Immunological Interest, National Institutes of Health, Bethesda, MD, (1987). Exemplary methods for producing  $V_H$  and  $V_L$  gene repertoires are described in PCT Application No. PCT/US 90/02836 (International Publication  
20 No. WO 90/14430).

In preferred embodiments, the preparation containing the total cellular mRNA is first enriched for the presence of  $V_H$  and/or  $V_L$  coding mRNA. Enrichment is typically accomplished by subjecting the total mRNA preparation or  
25 partially purified mRNA product thereof to a primer extension reaction employing a polynucleotide synthesis primer of the present invention.

According to another aspect of the present invention, a gene repertoire may be generated from one or a few  
30 nucleotide sequences by replicating those sequences under mutagenesis conditions so that a plurality of different nucleotide sequences or genes may be generated. Suitable mutagenesis conditions are known to those skilled in the art.

## 2. Preparation Of Polynucleotide Primers

The term "polynucleotide" as used herein in reference to primers, probes and nucleic acid fragments or segments to be synthesized by primer extension is defined as a molecule comprised of two or more deoxyribonucleotides or ribonucleotides, preferably more than 3. Its exact size will depend on many factors, which in turn depends on the ultimate conditions of use.

The term "primer" as used herein refers to a polynucleotide whether purified from a nucleic acid restriction digest or produced synthetically, which is capable of acting as a point of initiation of synthesis when placed under conditions in which synthesis of a primer extension product which is complementary to a nucleic acid strand is induced, i.e., in the presence of nucleotides and an agent for polymerization such as DNA polymerase, reverse transcriptase and the like, and at a suitable temperature and pH. The primer is preferably single stranded for maximum efficiency, but may alternatively be stranded. If double stranded, the primer is first treated to separate its strands before being used to prepare extension products. Preferably, the primer is a polydeoxyribonucleotide. The primer must be sufficiently long to prime the synthesis of extension products in the presence of the agents for polymerization. The exact lengths of the primers will depend on many factors, including temperature and the source of primer. For example, depending on the complexity of the target sequence, a polynucleotide primer typically contains 15 to 25 or more nucleotides, although it can contain fewer nucleotides. Short primer molecules generally require cooler temperatures to form sufficiently stable hybrid complexes with template.

The primers used herein are selected to be "substantially" complementary to the different strands of each specific sequence to be synthesized or amplified. This means that the primer must be sufficiently comple-



mentary to nonrandomly hybridize with its respective template strand. Therefore, the primer sequence need not reflect the exact sequence of the template. For example, a non-complementary nucleotide fragment can be attached to the 5' end of the primer, with the remainder of the primer sequence being substantially complementary to the strand. Such noncomplementary fragments typically code for an endonuclease restriction site. Alternatively, noncomplementary bases or longer sequences can be interspersed into the primer, provided the primer sequence has sufficient complementarity with the sequence of the strand to be synthesized to amplified to non-randomly hybridize therewith and thereby form an extension product under polynucleotide synthesizing conditions.

Primers of the present invention may also contain a DNA-dependent RNA polymerase promoter sequence or its complement. See for example, Krieg et al., Nucleic Acids Research, 12:7057-70 (1984); Studier et al., J. Mol. Biol., 189:113-130 (1986); and Molecular Cloning: A Laboratory Manual, Second Edition, Maniatis et al., eds., Cold Spring Harbor, NY (1989).

When a primer containing a DNA-dependent RNA polymerase promoter is used, the primer is hybridized to the polynucleotide strand to be amplified and the second polynucleotide strand of the DNA-dependent RNA polymerase promoter is completed using an inducing agent such as E. coli DNA polymerase I, or the Klenow fragment of E. coli DNA polymerase. The starting polynucleotide is amplified by alternating between the production of an RNA polynucleotide and DNA polynucleotide.

Primers may also contain a template sequence or replication initiation site for a RNA-directed polymerase. Typical RNA-directed RNA polymerase include the QB replicase described by Lizardi et al. Biotechnology, 6:1197-1202 (1988). RNA-directed polymerases produce large numbers of RNA strands from a small number of template RNA strands that contain a template sequence or replication

initiation site. These polymerases typically give a one million-fold amplification of the template strand, as has been described by Kramer et al., J. Mol. Biol., 89:7819-736 (1974).

5       The polynucleotide primers can be prepared using any suitable method, such as, for example, the phosphotriester on phosphodiester methods see Narang et al., Meth. Enzymol., 68:90, (1979); U.S. Patent No. 4,356,270; and Brown et al., Meth. Enzymol., 68:109, (1979).

10       The choice of a primer's nucleotide sequence depends on factors such as the distance on the nucleic acid from the region coding for the desired receptor, its hybridization site on the nucleic acid relative to any second primer to be used, the number of genes in the repertoire  
15 it is to hybridize to, and the like.

(a) Primers for Producing  $V_H$  and  $V_L$  DNA Homologs

$V_H$  and  $V_L$  gene repertoires can be separately prepared prior to their use in the methods of the present invention. Repertoire preparation is typically done by primer  
20 extension (or other in vitro amplification method), preferably by primer extension in a PCR format.

For example, to produce  $V_H$ -coding DNA homologs by primer extension, the nucleotide sequence of a primer is selected to hybridize with a plurality of immunoglobulin  
25 heavy chain genes at a site substantially adjacent to the  $V_H$ -coding region so that a nucleotide sequence coding for a functional (capable of finding) polypeptide is obtained. To hybridize to a plurality of different  $V_H$ -coding nucleic acid strands, the primer must be a substantial complement  
30 of a nucleotide sequence conserved among the different strands. Such sites include nucleotide sequences in the constant region, any of the variable region framework regions, preferably the third framework region, leader region, promoter region, J region and the like.

35       If the  $V_H$ -coding and  $V_L$ -coding DNA homologs are to be produced by polymerase chain reaction (PCR) amplification,

two primers must be used for each coding strand of nucleic acid to be amplified. The first primer becomes part of the nonsense (minus or complimentary) strand and hybridizes to a nucleotide sequence conserved among  $V_H$  (plus) strands within the repertoire. To produce  $V_H$  coding DNA homologs, first primers are therefore chosen to hybridize to (i.e. be complementary to) conserved regions within the J region, CH1 region, hinge region,  $C_H2$  region, or  $C_H3$  region of immunoglobulin genes and the like. To produce a  $V_L$  coding DNA homolog, first primers are chosen to hybridize with (i.e. be complementary to) a conserved region with the J region or constant region of immunoglobulin light chain genes and the like. Second primers become part of the coding (plus) strand and hybridize to a nucleotide sequence conserved among minus strands. To produce the  $V_H$ -coding DNA homologs, second primers are therefore chosen to hybridize with a conserved nucleotide sequence at the 5' end of the  $V_H$ -coding immunoglobulin gene such as in that area coding for the leader or first framework region. It should be noted that in the amplification of both  $V_H$ - and  $V_L$ -coding DNA homologs, the conserved 5' nucleotide sequence of the second primer can be complementary to a sequence exogenously added using terminal deoxynucleotidyl transferase as described by Loh et al., Science 243:217-220 (1989). One or both of the first and second primers can contain a nucleotide sequence defining an endonuclease recognition site. The site can be heterologous to the immunoglobulin gene being amplified and typically appears at or near the 5' end of the primer.

30 (b) Inside and Outside Primers

In one embodiment, the present invention utilizes a set of polynucleotides that form inside primers comprised of an upstream inside primer and a downstream inside primer. Each of the inside primers has a priming region located at the 3'-terminus of the primer. The priming region is typically the 3'-most (3'-terminal) 15 to 30

nucleotide bases. The 3'-terminal priming portion of each inside primer is capable of acting as a primer to catalyze nucleic acid synthesis, i.e., initiate a primer extension reaction off its 3' terminus. One or both of the inside primers is further characterized by the presence of a 5'-terminal (5'-most) non-priming portion, i.e., a region that does not participate in hybridization to repertoire template.

In fusion PCR, each inside primer works in combination with an outside primer to amplify a target nucleic acid sequence. The choice of PCR primer pairs for use in fusion PCR as described herein is governed by the same considerations as previously discussed for choosing PCR primer pairs useful in producing gen repertoires. That is, the primers have a nucleotide sequence that is complementary to a sequence conserved in the repertoire. Useful  $V_L$  and  $V_H$  inside priming sequences are shown in Tables 1 and 2, respectively, below.

Table 1

3' Priming Portions of Various Inside  $V_L$  Primers

Seq.

Id. No.

	(1) <sup>1</sup>	5' GTGATGACCCACTCTCC 3'
	(2)	5' GTGATGACCCAGTCTCCA 3'
25	(3)	5' GTTGTGACTCAGGAATCT 3'
	(4)	5' GTGTTGACGCAGCCGCCC 3'
	(5)	5' GTGCTCACCCAGTCTCCA 3'
	(6)	5' CAGATGACCCAGTCTCCA 3'
	(7)	5' GTGATGACCCAGACTCCA 3'
30	(8)	5' GTCATGACCCAGTCTCCA 3'
	(9)	5' TTGATGACCCAAACTCAA 3'
	(10)	5' GTGATAACCCAGGATGAA 3'

<sup>1</sup> Nucleotides sequences 1-10 are unique 5' primers for the amplification of kappa light chain variable regions.

Table 23' Priming Portions of Various Inside V<sub>H</sub> Primers

Seq.

Id. No.

5	(11) <sup>1</sup>	5' ACAAGATTTGGGCTC 3'
	(12) <sup>2</sup>	5' TGGGGTTTTGAGCTC 3'
	(13) <sup>3</sup>	5' GAGACAGTGACCGGGTTCCTTGGCCCCA 3'
	(14) <sup>4</sup>	5' TGGAATGGGCACATGCAG 3'
	(15) <sup>5</sup>	5' TTATCATTTACCCGGAGA 3'
10	(16) <sup>6</sup>	5' AACGGTAACAGTGGTGCCTTGGCCCCA 3'
	(17) <sup>7</sup>	5' ACAATCCCTGGGCACAAT 3'
	(18) <sup>8</sup>	5' CACCTTGGTGCTGCTGGC 3'
	(19) <sup>9</sup>	5' ACAACCACAATCCCTGGGCACAATTTT 3'
	(20) <sup>10</sup>	5' ACAATCCCTGGGCACAAT 3'
15	(21) <sup>11</sup>	5' GAGTTCACTAGTTGGGCACGGTGGGCA 3'
	<sup>1</sup>	Unique 3' primer for human IgG1, 2, 3, and 4 F.2d.
	<sup>2</sup>	Unique 3' primer for human V <sub>H</sub> amplification.
	<sup>3</sup>	3' primer for amplifying human heavy chain variable regions.
20	<sup>4</sup>	3' primer for amplifying the Fd region of mouse IgM.
	<sup>5</sup>	3' primer located in the CH3 region of human IgG1 to amplify the entire heavy chain.
	<sup>6</sup>	Unique 3' primer for amplification of mouse F <sub>v</sub> .
	<sup>7</sup>	Unique 3' primer for amplification of mouse IgG1 Fd.
25	<sup>8</sup>	Unique 3' primer for amplification of VH including part of the mouse gamma 1 first constant region.
	<sup>9</sup>	Unique 3' primer for amplification of VH including part of mouse gamma 1 first constant region and hinge region.
30	<sup>10</sup>	3' primer for amplifying mouse Fd including part of the mouse IgG first constant region and part of the hinge region.
	<sup>11</sup>	3' primer for amplifying human IgG1 Fd including part of the human IgG first constant region and part of the hinge region including the two cysteines which create the disulfide bridge for producing Fab'2 (the primer corresponds to Kabat number 241QQ to 247).
35		

A preferred set of inside primers used herein has primers with complementary 5'-terminal non-priming regions, the complementary strands of which are capable of hybridizing to each other to form a duplex with 3' overhangs. The duplex encodes all or part of a double stranded cistronic bridge. That is, if the 3' overhangs of the duplex are filled in with complementary bases so as to define a double stranded DNA extending from the 3'-terminus of one of the inside primers to the 3'-terminus of the other of the inside primers, that double stranded DNA segment forms a sequence of nucleotides that operatively links the upstream and downstream cistrons for polycistronic expression. Thus, while each of the inside primers in a set contains only a portion of the sequence information necessary to form the double stranded cistronic bridge, the two inside primers in combination encode both the plus and minus strands of all or part of the bridge.

For example, one inside upstream primer can have a sequence that forms a portion of the plus strand of the bridge, and the other inside primer encodes the sequence, through complementarity, of the downstream portion of the plus strand.

In a preferred embodiment, the plus strand of the cistronic bridge contains, in the translational reading frame and from an upstream position to a downstream position, sequences coding for (i) at least one stop codon, preferably two, in the same reading frame as the upstream cistron, (ii) a ribosome binding site, and (iii) a polypeptide leader, the translation initiation codon of which is in the same reading frame as the downstream cistron. The stop codon is present to terminate translation of the upstream cistron. The ribosome binding site is present to initiate translation of the downstream cistron from the polycistronic mRNA.

The predicted amino acid residue sequences of two *pelB* gene product variants from Erwinia Carotova are shown

**SUBSTITUTE SHEET**

in Table 3. Lei, et al., supra. Amino Acid residue sequences for other leaders from E. coli useful in this invention are also listed in Table 3. Oliver, In Neidhart, F. C. (ed.), Escherichia coli and Salmonella  
 5 Typhimurium, American Society for Microbiology, Washington, D. C., 1:56-69 (1987). These regions for the heavy chain are contained in the modified ImmunoZAP H expression vector. Mullinax, et al., Proc. Natl. Acad. Sci., USA, 87:8095-8099 (1990).

10 Table 3  
Leader Sequences

Seq.

<u>Id. No.</u>	<u>Type</u>	<u>Amino Acid Residue Sequence</u>
(22)	pelB <sup>1</sup>	MetLysTyrLeuLeuProThrAlaAlaAlaGlyLeuLeu
15		LeuLeuAlaAlaGlnProAlaGlnProAlaMetAla
(23)	pelB <sup>2</sup>	MetLysSerLeuIleThrProIleAlaAlaGlyLeuLeu
		LeuAlaPheSerGlnTyrSerLeuAla
(24)	MalE <sup>3</sup>	MetLysIleLysThrGlyAlaArgIleLeuAlaLeuSer
		AlaLeuThrThrMetMetPheSerAlaSerAlaLeuAla
20		LysIle
(25)	OmpF <sup>3</sup>	MetMetLysArgAsnIleLeuAlaValIleValProAla
		LeuLeuValAlaGlyThrAlaAsnAlaAlaGlu
(26)	PhoA <sup>3</sup>	MetLysGlnSerThrIleAlaLeuAlaLeuLeuProLeu
		LeuPheThrProValThrLysAlaArgThr
25	(27)	Bla <sup>3</sup> MetSerIleGlnHisPheArgValAlaLeuIleProPhe
		PheAlaAlaPheCysLeuProValPheAlaHisPro
(28)	LamB <sup>3</sup>	MetMetIleThrLeuArgLysLeuProLeuAlaValAla
		ValAlaAlaGlyValMetSerAlaGlnAlaMetAlaVal
		Asp
30	(29)	Lpp <sup>3</sup> MetLysAlaThrLysLeuValLeuGlyAlaValIleLeu
		GlySerThrLeuLeuAlaGlyCysSer

1 pelB from Erwinia carotovora gene

2 pelB from Erwinia carotovora EC 16 gene

35 3 leader sequences from E. coli

To achieve high levels of gene expression in E. coli, it is necessary to use not only strong promoters to generate large quantities of mRNA, but also ribosome binding sites to ensure that the mRNA is efficiently translated.

5 In E. coli, the ribosome binding site includes an initiation codon (AUG) and a sequence 3- nucleotides long located 3 11 nucleotides upstream from the initiation codon [Shine et al., Nature, 254:34 (1975)]. The sequence, AGGAGGU, which is called the Shine-Dalgarno (SD) sequence, is complementary to the 3' end of E. coli 16S mRNA. Binding of the ribosome to mRNA and the sequence at the 3' end of the mRNA can be affected by several factors:

(i) The degree of complementarity between the SD sequence and 3' end of the 16S tRNA.

15 (ii) The spacing and possibly the DNA sequence lying between the SD sequence and the AUG [Roberts et al., Proc. Natl. Acad. Sci. USA, 76:760 (1979A); Roberts et al., Proc. Natl. Acad. Sci. USA, 76:5596 (1979B); Guarente et al., Science, 209:1428 (1980); and Guarente et al., Cell, 20:543 (1980).] Optimization is achieved by measuring the level of expression of genes in plasmids in which this spacing is systematically altered. Comparison of different mRNAs shows that there are statistically preferred sequences from positions -20 to +13 (where the A of the AUG is position 0) [Gold et al., Annu. Rev. Microbiol., 25 35:365 (1981)]. Leader sequences have been shown to influence translation dramatically (Roberts et al. 1979 a, b supra).

(iii) The nucleotide sequence following the AUG, 30 which affects ribosome binding [Taniguchi et al., J. Mol. Biol., 118:533 (1978)].

Useful ribosome binding sites are shown in Table 4 below.



Table 4Ribosome Binding Sites\*

Seq.

Id. No.

5	1.	(30)	5' AAUCUUGGAGGCUUUUUU <u>AUG</u> GUUCGUUCU
	2.	(31)	5' UAACUAAGGAUGAAAUGCA <u>UG</u> UCUAAGACA
	3.	(32)	5' UCCUAGGAGGUUUGACCU <u>AUG</u> CGAGCUUUU
	4.	(33)	5' AUGUACUAAGGAGGUUGU <u>AUG</u> GAACAACGC

\* Sequences of initiation regions for protein  
10 synthesis in four phage mRNA molecules are underlined.

AUG = initiation codon (double underlined)

1. = Phage  $\phi$ X174 gene-A protein
2. = Phage Q $\beta$  replicase
3. = Phage R17 gene-A protein
- 15 4. = Phage lambda gene-cro protein

It is preferred that the complementary (overlapping)  
region of the inside primers and the priming portion of  
the inside primers have about the same denaturation  
temperature, Td. The Td of a sequence can be estimated by  
20 the following formula:  $Td = 4(C+G) + 2(A+T)$ , where C, G,  
A and T represent the respective number of cytosine,  
guanine, adenine and thymine bases in the sequence. A Td  
for the above-identified hybridizing region of about 45-  
55°C, preferably about 50°C, is preferred. Typically,  
25 overlapping regions in the range of about 15 to 20  
nucleotides works well in conjunction with the priming  
regions in the range of 15-30 nucleotides.

The set of outside primers forms the termini of the  
dicistronic DNA molecule. The set of outside primers  
30 comprises an upstream outside primer and a downstream  
outside primer. The outside primers each comprise a 3'-  
terminal priming portion, and preferably a portion that  
defines an endonuclease restriction site. When present,  
the restriction site-defining portion is typically located  
35 in a 5'-terminal non-priming portion of the outside  
primer. The restriction site defined by the upstream  
outside primer is typically chosen to be one recognized by

**SUBSTITUTE SHEET**

a restriction enzyme that does not recognize the restriction site defined by the downstream outside primer, the objective being to be able to produce a dicistronic DNA having cohesive termini that are non-complementary to each other and thus allow directional insertion into a vector.

Useful outside primer sequences are shown in Tables 5 and 6 below.

Table 5  
10 Outside V<sub>H</sub> Primers

Seq.		<u>Id. No.</u>	
	(34) <sup>1</sup>	5'	AGGTCCAGCTGCTCGAGTCTGG3'
	(35)	5'	AGGTCCAGCTGCTCGAGTCAGG3'
15	(36)	5'	AGGTCCAGCTTCTCGAGTCTGG3'
	(37)	5'	AGGTCCAGCTTCTCGAGTCAGG3'
	(38)	5'	AGGTCCAAGTCTCGAGTCTGG3'
	(39)	5'	AGGTCCAAGTCTCGAGTCAGG3'
	(40)	5'	AGGTCCAAGTCTCGAGTCTGG3'
20	(41)	5'	AGGTCCAAGTCTCGAGTCAGG3'
	(42) <sup>2</sup>	5'	AGGTGCAGCTGCTCGAGTCTGG3'
	(43)	5'	AGGTGCAGCTGCTCGAGTCGGG3'
	(44)	5'	AGGTGCAAGTCTCGAGTCTGG3'
	(45)	5'	AGGTGCAAGTCTCGAGTCGGG3'
25	<sup>1</sup> Nucleotide sequences 21-28 are unique 5' primers for the amplification of mouse V <sub>H</sub> genes.		
	<sup>2</sup> Nucleotide sequences 29-32 are unique 5' primers for amplification of nucleic acids coding for human variable regions.		

30 Table 6  
Outside V<sub>L</sub> Primers

Seq.		<u>Id. No.</u>	
	(46) <sup>1</sup>	5'	ACGTCTAGATTCCACCTTGGTCCC 3'
35	(47) <sup>2</sup>	5'	TCCTTCTAGATTACTAACACTCTCCCCTGTTGAA 3'

- (48)<sup>3</sup> 5' GCATTCTAGACTATTAACATTCTGTAGGGGC 3'
- (49)<sup>4</sup> 5' GCAGCATTCTAGAGTTTCAGCTCCAGCTTGCC 3'
- (50)<sup>5</sup> 5' CCGCCGTCTAGAACACTCATTCTGTGAAGCT 3'
- (51)<sup>6</sup> 5' CCGCCGTCTAGAACATTCTGCAGGAGACAGACT 3'
- 5 (52)<sup>7</sup> 5' GCGCCGTCTAGAATTAACACTCATTCTGTGAA 3'
- (53)<sup>8</sup> 5' GCCGCTCTAGAACACTCATTCTGTGAA 3'
- (54)<sup>9</sup> 5' TCCTTCTAGATTACTAACACTCTCCCCTGTGAA 3'
- (55)<sup>10</sup> 5' GCATTCTAGACTATTATGAACATTCTGTAGGGGC 3'
- 1 3' primer for amplifying human kappa chain variable  
10 regions.
- 2 3' primer in human kappa light chain constant region.
- 3 3' primer in human lambda light chain constant  
region.
- 4 Unique 3' primer for amplification of kappa light  
15 chain variable regions.
- 5 Unique 3' primer for mouse kappa light chain  
amplification including the constant region.
- 6 Unique 3' primer for mouse lambda light chain  
amplification including the constant region.
- 20 7 Unique 3' primer for amplification of kappa light  
chain.
- 8 Unique 3' primer for amplification of mouse kappa  
light chain.
- 9 Unique 3' primer for kappa V<sub>L</sub> amplification.
- 25 10 Unique 3' primer for human, mouse and rabbit lambda  
V<sub>L</sub> amplification.

### 3. Preparing a Gene Library

The strategy used for cloning, i.e., substantially reproducing the V<sub>H</sub> and/or V<sub>L</sub> genes contained within the  
30 isolated repertoire will depend, as is well known in the art, on the type, complexity, and purity of the nucleic acids making up the repertoire. Other factors include whether or not the genes are contained in one or a plurality of repertoires or populations and whether or not  
35 they are to be amplified and/or mutagenized.

a. Preparing  $V_H$  and  $V_L$  libraries

In one strategy, the object is to clone the  $V_H$ - and/or  $V_L$ -coding genes from a repertoire comprised of polynucleotide coding strands, such as mRNA and/or the sense strand of genomic DNA. If the repertoire is in the form of double stranded genomic DNA, it is usually first denatured, typically by melting, into single strands. The repertoire is subjected to a first primer extension reaction by treating (contacting) the repertoire with a first polynucleotide synthesis primer having a preselected nucleotide sequence. The first primer is capable of initiating the first primer extension reaction by hybridizing to a nucleotide sequence, preferably at least about 10 nucleotides in length and more preferably at least about 20 nucleotides in length, conserved within the repertoire. The first primer is sometimes referred to herein as the "sense primer" because it hybridizes to the coding or sense strand of a nucleic acid. In addition, the second primer is sometimes referred to herein as the "anti-sense primer" because it hybridizes to a non-coding or anti-sense strand to a nucleic acid, i.e., a strand complementary to a coding strand.

The PCR reaction is performed by mixing the PCR pair, preferably a predetermined amount thereof, with the nucleic acids of the repertoire, preferably a predetermined amount thereof, in a PCR buffer to form a first PCR admixture. The admixture is maintained under polynucleotide synthesizing conditions for a time period, which is typically predetermined, sufficient for the formation of a PCR reaction product, thereby producing a gene library containing a plurality of different  $V_H$ - and/or  $V_L$ -coding DNA homologs.

A plurality of first primer and/or a plurality of second primers can be used in each amplification, e.g., one species of first primer can be paired with a number of second primers to form several different primer pairs. Alternatively, an individual pair of first and second

primers can be used. In any case, the amplification products of amplifications using the same or different combinations of first and second primers can be combined to increase the diversity of the gene library.

5 In another strategy, the object is to clone the  $V_H$ - and/or  $V_L$ -coding gene from a repertoire by providing a polynucleotide complement of the repertoire, such as the anti-sense strand of genomic dsDNA or the polynucleotide produced by subjecting mRNA to a reverse transcriptase  
10 reaction. Methods for producing such complements are well known in the art. The complement is subjected to a primer extension reaction similar to the above-described second primer extension reaction, i.e., a primer extension reaction using a polynucleotide synthesis primer capable  
15 to hybridizing to a nucleotide sequence conserved among a plurality of different  $V_H$ -coding gene complements.

The primer extension reaction is performed using any suitable method. Generally it occurs in a buffered aqueous solution, preferably at a pH of 7-9, most preferably  
20 about 8. Preferably, a molar excess (for genomic nucleic acid, usually about  $10^6$ :1 primer:template) of the primer is admixed to the buffer containing the template strand. A large molar excess is preferred to improve the efficiency of the process.

25 The deoxyribonucleotide triphosphates dATP, dCTP, dGTP, and dTTP are also admixed to the primer extension (polynucleotide synthesis) reaction admixture in adequate amounts and the resulting solution is heated to about  $90^{\circ}\text{C}$ - $100^{\circ}\text{C}$  for about 1 to 10 minutes, preferably from 1  
30 to 4 minutes. After this heating period the solution is allowed to cool to room temperature, which is preferable for primer hybridization. To the cooled mixture is added an appropriate agent for inducing or catalyzing the primer extension reaction, and the reaction is allowed to occur  
35 under conditions known in the art. The synthesis reaction may occur at from room temperature up to a temperature above which the inducing agent no longer functions effi-

**SUBSTITUTE SHEET**

ciently. Thus, for example, if DNA polymerase is used as inducing agent, the temperature is generally no greater than about 40°C.

The inducing agent may be any compound or system which will function to accomplish the synthesis of primer extension products, including enzymes. Suitable enzymes for this purpose include, for example, E. coli, DNA polymerase I, Klenow fragment of E. coli DNA polymerase I, T4 DNA polymerase, other available DNA polymerases, reverse transcriptase, and other enzymes, including heat-stable enzymes, which will facilitate combination of the nucleotides in the proper manner to form the primer extension products which are complementary to each nucleic acid strand. Generally, the synthesis will be initiated at the 3' end of each primer and proceed in the 5' direction along the template strand, until synthesis terminates, producing molecules of different lengths. There may be inducing agents, however, which initiate synthesis at the 5' end and proceed in the above direction, using the same process as described above.

The inducing agent also may be a compound or system which will function to accomplish the synthesis of RNA primer extension products, including enzymes. In preferred embodiments, the inducing agent may be a DNA-dependent RNA polymerase such as T7 RNA polymerase, T3 RNA polymerase or SP6 RNA polymerase. These polymerases produce a complementary RNA polynucleotide. The high turnover rate of the RNA polymerase amplifies the starting polynucleotide as has been described by Chamberlin et al., The Enzymes, ed. P. Boyer, PP. 87-108, Academic Press, New York (1982). Another advantage of T7 RNA polymerase is that mutations can be introduced into the polynucleotide synthesis by replacing a portion of cDNA with one or more mutagenic oligodeoxynucleotides (polynucleotides) and transcribing the partially-mismatched template directly as has been previously described by Joyce et al., Nucleic Acid Research, 17:711-722 (1989). Amplification systems

based on transcription have been described by Gingeras et al., in PCR Protocols, A Guide to Methods and Applications, PP. 245-252, Academic Press, Inc., San Diego, CA (1990).

5        If the inducing agent is a DNA-dependent RNA polymerase and therefore incorporates ribonucleotide triphosphates, sufficient amounts of ATP, CTP, GTP and UTP are admixed to the primer extension reaction admixture and the resulting solution is treated as described above.

10       The newly synthesized strand and its complementary nucleic acid strand form a double-stranded molecule which can be used in the succeeding steps of the process.

      The first and/or second primer extension reaction discussed above can advantageously be used to incorporate  
15       into the multimeric polypeptide a preselected epitope useful in immunologically detecting and/or isolating a multimeric polypeptide. This is accomplished by utilizing a first and/or second polynucleotide synthesis primer or expression vector to incorporate a predetermined amino  
20       acid residue sequence into the amino acid residue sequence of the receptor.

      After producing  $V_H$ - and/or  $V_L$ -coding DNA homologs for a plurality of different  $V_H$ - and/or  $V_L$ -coding genes within the repertoire, the homologs are typically amplified.  
25       While the  $V_H$  and/or  $V_L$ -coding DNA homologs can be amplified by classic techniques such as incorporation into an autonomously replicating vector, it is preferred to first amplify the DNA homologs by subjecting them to a polymerase chain reaction (PCR) prior to inserting them into a  
30       vector. In fact, in preferred strategies, the first and/or second primer extension reactions used to produce the gene library are the first and second primer extension reactions in a polymerase chain reaction.

      PCR is typically carried out by cycling i.e.,  
35       simultaneously performing in one admixture, the above described first and second primer extension reactions, each cycle comprising polynucleotide synthesis followed by

denaturation of the double stranded polynucleotides formed. Methods and systems for amplifying a DNA homolog are described in U.S. Patents No. 4,683,195 and No. 4,683,202, both to Mullis et al. Preferably, PCR is  
5 carried out by thermocycling i.e., repeatedly increasing and decreasing the temperature of a PCR reaction admixture within a temperature range whose lower limit is about 10°C to about 50°C and whose upper limit is about 90°C to about 100°C. The increasing and decreasing can be continuous,  
10 but is preferably phasic with time periods of relative temperature stability at each of temperatures favoring polynucleotide synthesis, denaturation and hybridization.

In preferred embodiments only one pair of first and second primers is used per amplification reaction. The  
15 amplification reaction products obtained from a plurality of different amplifications, each using a plurality of different primer pairs, are then combined.

However, the present invention also contemplated DNA homolog production via co-amplification (using two pairs  
20 of primers), and multiplex amplification (using up to about 8, 9 or 10 primer pairs).

The  $V_H$ - and  $V_L$ -coding DNA homologs produced by PCR amplification are typically in double-stranded form and have contiguous or adjacent to each of their termini a  
25 nucleotide sequence defining an endonuclease restriction site. Digestion of the  $V_H$ - and  $V_L$ -coding DNA homologs having restriction sites at or near their termini with one or more appropriate endonucleases results in the production of homologs having cohesive termini of predetermined  
30 specificity.

In preferred embodiments, the PCR process is used not only to amplify the  $V_H$ - and/or  $V_L$ -coding DNA homologs of the library, but also to induce mutations within the library and thereby provide a library having a greater  
35 heterogeneity. First, it should be noted that the PCR processes itself is inherently mutagenic due to a variety of factors well known in the art. Second, in addition to



the mutation inducing variations described in the above referenced U.S. Patent No. 4,683,195, other mutation inducing PCR variations can be employed. For example, the PCR reaction admixture, i.e., the combined first and second primer extension reaction admixtures, can be formed with different amounts of one or more of the nucleotides to be incorporated into the extension product. Under such conditions, the PCR reaction proceeds to produce nucleotide substitutions within the extension product as a result of the scarcity of a particular base. Similarly, approximately equal molar amounts of the nucleotides can be incorporated into the initial PCR reaction admixture in an amount to efficiently perform X number of cycles, and then cycling the admixture through a number of cycles in excess of X, such as, for instance, 2X. Alternatively, mutations can be induced during the PCR reaction by incorporating into the reaction admixture nucleotide derivatives such as inosine, not normally found in the nucleic acids of the repertoire being amplified. During subsequent in vivo amplification, the nucleotide derivative will be replaced with a substitute nucleotide thereby inducing a point mutation.

b. Preparing a Dicistronic DNA molecule Library

In one embodiment, a library of dicistronic DNA molecules containing upstream and downstream cistrons operatively linked by a cistronic bridge can be produced by the following steps:

(a) Subjecting a repertoire of first polypeptide genes (e.g.,  $V_H$ -coding genes), to PCR amplification using first outside and first inside primers, i.e., a first PCR primer pair, to form a first primary PCR product.

(b) Subjecting a repertoire of second polypeptide genes (e.g.,  $V_L$ -coding genes) to PCR amplification using second outside and second inside primers, i.e., a second PCR primer pair, to form a second primary PCR product.

(c) Hybridizing the first and second primary PCR products to form internally (self) primed duplexes, i.e., duplexes having 3'-hybridized and 5'-overhanging termini.

(d) Subjecting the internally-primed duplexes to  
5 primer extension reaction conditions to form double stranded duplexes having substantially blunt, preferably blunt, termini and a dicistronic strand containing the upstream and downstream cistrons linked by a cistronic bridge encoded by the inside primers. By "substantially  
10 blunt" is meant having no more than about one or two overhanging nucleotides. (Substantially blunt double stranded DNA is sometimes produced by primer overextension by Taq polymerase, usually by the addition of one or two terminal adenine residues.)

15 The  $V_H$ - and  $V_L$ -coding gene repertoires are comprised of polynucleotide coding strands, such as mRNA and/or the sense strand of genomic DNA. If the repertoire is in the form of double stranded genomic DNA, it is usually first denatured, typically by melting, into single strands. A  
20 repertoire is subjected to a PCR reaction as described in Section 3a hereinabove.

In preferred embodiments the ratio of gene molecules and their respective primers is as follows: about  $1 \times 10^3$   $V_H$  gene molecules to about  $1 \times 10^8$  outside  $V_H$  gene molecules  
25 to about  $1 \times 10^8$  outside  $V_H$  primer molecules, about  $1 \times 10^3$   $V_H$  gene molecules, to about  $1 \times 10^7$  inside  $V_H$  gene primer molecules, about  $1 \times 10^3$   $V_L$  gene molecules to about  $1 \times 10^8$  outside  $V_L$  gene primer molecules, about  $1 \times 10^4$   $V_L$  gene molecules to about  $1 \times 10^7$   $V_L$  gene primer molecules. In  
30 more preferred embodiments,  $10^4$  outside  $V_H$  gene primer molecules and  $10^3$  inside  $V_H$  gene primer molecules are used for every  $V_H$  gene molecule present in the PCR admixture. Similarly,  $10^4$  outside  $V_L$  gene primer molecules and  $10^3$   $V_L$  gene molecule present in the PCR admixture. Thus, there  
35 is typically a 10 fold molar excess of outside primer to inside primer.

In the fusion PCR reaction, the gene repertoires are admixed with outside and inside primers, the outside primers being present in excess relative to the inside primers. The initial PCR thermocycles produce intermediate products having complementary termini from each of the first and second gene repertoires. That is, the end of one strand from one primary PCR product is capable of hybridizing with the complementary end from the other primary PCR product. The strands having the overlap at their 3' ends can act as primers for one another, i.e., from an internally primed duplex, and be extended by the polymerase to form the full length final product. The final product is then amplified by the set of outside primers, which act as a third PCR pair when the inside primers have been exhausted, to form a secondary PCR product. Typically the molar ratio of outside primers to inside primers is such that the inside primers are effectively exhausted within about 2 to about 12, preferably about 5, 6 or 7 thermocycles.

The PCR buffer also contains the deoxyribonucleotide triphosphates dATP, dCTP, dGTP, and a polymerase, typically thermostable, all in adequate amounts for primer extension (polynucleotide synthesis) reaction. The resulting solution (PCR admixture) is heated to about 90°C - 100°C for about 1 to 10 minutes, preferably from 1 to 4 minutes. After this heating period the solution is allowed to cool to 54°C, which is preferably for primer hybridization. The synthesis reaction may occur at from room temperature up to a temperature above which the polymerase (inducing agent) no longer functions efficiently. Thus, for example, if DNA polymerase is used as inducing agent, the temperature is generally no greater than about 40°C. An exemplary PCR buffer comprises the following: 50 mM KCl; 10 mM Tris-HCl; pH 8.3; 1.5 mM MgCl<sub>2</sub>; 0.001% (wt/vol) gelatin, 200 μM dATP; 200 μM dTTP; 200 μM dCTP; 200 μM dGTP; and 2.5 units Thermus aquaticus DNA poly-

merase I (U.S. Patent No. 4,889,818) per 100 microliters of buffer.

After producing operatively linked  $V_H$ - and  $V_L$ -coding DNA homologs for a plurality of different  $V_H$ - and  $V_L$ - coding genes within the repertoires, the dicistronic DNA molecules are typically further amplified. While the dicistronic DNA molecules can be amplified by classic techniques such as incorporation into an autonomously replicating vector, it is preferred to first amplify the molecules by subjecting them to a polymerase chain reaction (PCR) prior to inserting them into a vector. In fact, in preferred strategies, the first and second PCR reactions are performed in the same admixture that is subject to a multiplicity of PCR thermocycles where the outside primers are in molar excess. Preferably the number of PCR thermocycles is at least  $n+5$ , wherein  $n$  is the number of PCR thermocycles necessary to decrease by a factor of 10, and preferably exhaust, the number of inside primers by consumption in the formation of inside primer-primed products.

A diverse library of dicistronic DNA molecules having upstream and downstream cistrons can also be produced by combining, in a PCR buffer, double stranded  $V_H$  and  $V_L$  repertoires,  $V_H$  and  $V_L$  outside primers, and an inside primer having a 3'-terminal priming portion, a cistronic bridge coding portion, and a 5'-terminal inside primer-template (primer-coding) portion. The 3'-terminal priming portion has a nucleotide base sequence complementary to a portion of the primer extension product of one of the outside primers. The 5'-terminal primer-template portion has a nucleotide base sequence homologous (identical) to a portion of the primer extension product of the other of the outside primers. That is, the linking primer has terminal sequences homologous to sequences in both repertoires. The cistronic bridge coding portion codes for, either directly or through complementarily, at least one stop codon in the same reading frame as the upstream

cistron and sequences for the expression of the downstream cistron.

The dicistronic DNA molecules containing operatively linked  $V_H$ - and  $V_L$ -coding DNA homologs produced by PCR  
5 amplification are typically in double-stranded in form and may have contiguous or adjacent to each of their termini a nucleotide sequence defining an endonuclease restriction site. Digestion of the dicistronic DNA molecules having restriction sites at or near their termini with one or more  
10 appropriate endonucleases results in the production of DNA molecules having cohesive termini of predetermined specificity.

When individual PCR admixtures contain diverse gene repertoires the present invention produces many non-  
15 naturally occurring antibodies, i.e., combinations of  $V_H$  and  $V_L$  in a heterodimer. To take advantage of the mammalian immune system's capacity to select  $V_H$  and  $V_L$  combinations, the present invention also contemplates using fusion PCR to operatively link, and thereby recover,  
20 naturally occurring  $V_H$  and  $V_L$  combinations.

In certain preferred embodiments, a fusion PCR method is performed on repertoires comprising a plurality of substantially isolated cells containing genes coding for a heterodimeric receptor. For example, a plurality of PCR  
25 admixtures is formed, each of which contains (i) a sample of substantially isolated B lymphocytes from a mammal producing antibody molecules against a preselected antigen, (ii) a PCR buffer, and (iii) either the previously described  $V_H$  and  $V_L$  PCR primer pairs or the set of outside  
30  $V_H$  and  $V_L$  PCR primers in combination with the linking primer(s), also as previously described. The plurality of PCR admixtures is then subjected to a multiplicity of PCR thermocycles as described herein.

By "substantially isolated" is meant a sample  
35 containing less than about 100 target cells, such as B lymphocytes, T cells, and the like. In preferred embodiments, the plurality of PCR admixtures contain only about

one cell. The cells are typically obtained from an individual mammal whose serum contains antibody molecules against the preselcted antigen. The collected cells are typically seeded, usually at densities in the range of 0.5 to 100 cells per unit volume, into a plurality of individual PCR vessels, such as microtiter plate wells and the like. Usually, the plurality of PCR admixtures is in the range of 800 to 1200, and preferably is about 1000, separate admixtures.

Typically, fewer cells are needed in each PCR admixture where the cells are obtained from individuals expressing a high serum antibody titer against the preselected antigen. For example, where B lymphocytes are obtained from an individual having a frequency of circulating B cells producing the antibody molecules of preselected specificity of 1/3000, each of about 800 to 1200 individual PCR admixtures need only contain about one B lymphocyte to result in isolation of the desired antibody. Where the circulating B cell frequency is in the range of 1/500,000, a density of about 100 cells per PCR admixture in each of about 800 to 1200 individual PCR admixtures will be needed before the process will result in isolation of the desired antibody.

In preferred embodiments, the PCR process is used not only to produce a library of dicistronic DNA molecules, but also to induce mutations within the library or to create diversity from a single parental clone and thereby provide a library having a greater heterogeneity as noted in Section 3a hereinabove.

#### 4. Expression

##### A. Expressing the $V_H$ and/or $V_L$ DNA Homologs.

The  $V_H$ - and/or  $V_L$ -coding DNA homologs contained within the library produced by the above-described method can be operatively linked to a vector for amplification and/or expression.

The choice of vector to which a  $V_H$ - and/or  $V_L$ -coding DNA homolog is operatively linked depends directly, as is well known in the art, on the functional properties desired, e.g., replication or protein expression, and the host cell to be transformed, these being limitations inherent in the art of constructing recombinant DNA molecules. In preferred embodiments, the vector utilized includes a procaryotic replicon i.e., a DNA sequence having the ability to direct autonomous replication and maintenance of the recombinant DNA molecule extra chromo-  
somally in a procaryotic host cell, such as a bacterial host cell, transformed therewith. Such replicons are well known in the art. In addition, those embodiments that include a procaryotic replicon also include a gene whose expression confers a selective advantage, such as drug resistance, to a bacterial host transformed therewith. Typical bacterial drug resistance genes are those that confer resistance to ampicillin or tetracycline.

Those vectors that include a procaryotic replicon can also include a procaryotic promoter capable of directing the expression (transcription and translation) of the  $V_H$ - and/or  $V_L$ -coding homologs in a bacterial host cell, such as E. coli transformed therewith. A promoter is an expression control element formed by a DNA sequence that permits binding of RNA polymerase and transcription to occur. Promoter sequences compatible with bacterial hosts are typically provided in plasmid vectors containing convenience restriction sites for insertion of a DNA segment of the present invention. Typical of such vector plasmids are pUC8, pUC9, pBR322, and pBR329 available from BioRad Laboratories, (Richmond, CA) and pPL and pKK223 available from Pharmacia, (Piscataway, NJ).

Promoters contain two highly conserved regions, one located about 10 bp (-10 region on Priberrow box) and the other about 35 bp (-35 region) upstream from the point at which transcription starts. These two regions typically determine promoter strength. In addition, the number of

nucleotides atht separate the conserved sequences is important for efficient promoter function. For example, 16 to 19 nucleotides typically separate the -10 and -35 regions, and changes in that psacing can change the efficiency of a promoter.

Promoters useful in this invention include Ptac  $\phi$  1.1A,  $\phi$  1.1B and  $\phi$  10, which are recognized by T7 polymerase. See U.S. Patent No. 4,946,786. Useful regulatable promoters include the E. coli lac promoter described in U.S. Patent No. 4,936,786 and the promoters for the temperature sensitive genes in U.S. Patent No. 4,806,471. See also U.S. Patent No. 4,711,845.

Expression vectors compatible with eukaryotic cells, preferably those compatible with vertebrate cells, can also be used. Eukaryotic cell expression vectors are well known in the art and are available from several commercial sources. Typically, such vectors are provided containing convenient restriction sites for insertion of the desired DNA homologue. Typical of such vectors are pSV<sub>1</sub> and pKSV-10 (Pharmacia), pBPV-1/PML2d (International Biotechnologies, Inc.), and pTDT1 (ATCC, No. 31255).

In preferred embodiments, the eukaryotic cell expression vectors used include a selection marker that is effective in an eukaryotic cell, preferably a drug resistance selection marker. A preferred drug resistance marker is the gene whose expression results in neomycin resistance, i.e., the neomycin phosphotransferase (neo) gene. Southern et al., J. Mol. Appl. Genet., 1:327-341 (1982).

The use of retroviral expression vectors to express the genes of the V<sub>H</sub> and/or V<sub>L</sub>-coding DNA homologs is also contemplated. As used herein, the term "retroviral expression vector" refers to a DNA molecule that includes a promoter sequences derived from the long terminal repeat (LTR) region of a retrovirus genome.

In preferred embodiments, the expression vector is typically a retroviral expression vector that is preferably replication-incompetent in eukaryotic cells. The



construction and use of retroviral vectors has been described by Sorge et al., Mol. Cel. Biol., 41730-1737 (1984).

5 A variety of methods have been developed to operatively link DNA to vectors via complementary cohesive termini. For instance, complementary cohesive termini can be engineered into the  $V_H$ - and/or  $V_L$ -coding DNA homologs during the primer extension reaction by use of an appropriately designed polynucleotide synthesis primer, as  
10 previously discussed. The vector, and DNA homolog if necessary, is cleaved with a restriction endonuclease to produce termini complementary to those of the DNA homolog. The complementary cohesive termini of the vector and the DNA homolog are then operatively linked (ligated) to  
15 produce a unitary double stranded DNA molecule.

In preferred embodiments, the  $V_H$ -coding and  $V_L$ -coding DNA homologs of diverse libraries are randomly combined in vitro for polycistronic expression from individual vectors. That is, a diverse population of double stranded  
20 DNA expression vectors is produced wherein each vector expresses, under the control of a single promoter, one  $V_H$ -coding DNA homolog and one  $V_L$ -coding DNA homolog, the diversity of the population being the result of different  $V_H$ - and  $V_L$ -coding DNA homolog combinations.

25 Random combination in vitro can be accomplished using two expression vectors distinguished from one another by the location on each of a restriction site common to both. Preferably the vectors are linear double stranded DNA, such as a Lambda Zap derived vector as described herein.  
30 In the first vector, the site is located between a promoter and a polylinker, i.e., 5' terminal (upstream relative to the direction of expression) to the polylinker by 3' terminal (downstream relative to the direction of expression). In the second vector, the polylinker is located  
35 between a promoter and the restriction site, i.e., the restriction site is located 3' terminal to the polylinker, and polylinker is located 3' terminal to the promoter.

**SUBSTITUTE SHEET**

In preferred embodiments, each of the vectors defines a nucleotide sequence coding for a ribosome binding and a leader, the sequence being located between the promoter and the polylinker, but downstream (3' terminal) from the shared restriction site if that site is between the promoter and polylinker. Also preferred are vectors containing a stop codon downstream from the polylinker, but upstream from any shared restriction site if that site is downstream from the polylinker. The first and/or second vector can also define a nucleotide sequence coding for a peptide tag. The tag sequence is typically located downstream from the polylinker but upstream from any stop codon that may be present.

In preferred embodiments, the vectors contain selectable markers such that the presence of a portion of that vector, i.e. a particular lambda arm, can be selected for or selected against. Typical selectable markers are well known to those skilled in the art. Examples of such markers are antibiotic resistance genes, genetically selectable markers, mutation suppressors such as amber suppressors and the like. The selectable markers are typically located upstream of the promoter and/or downstream of the second restriction site. In preferred embodiments, one selectable marker is located upstream of the promoter on the first vector containing the  $V_H$ -coding DNA homologs. A second selectable marker is located downstream of the second restriction site on the vector containing the  $V_L$ -coding DNA homologs. This second selectable marker may be the same or different from the first as long as when the  $V_H$ -coding vectors and the  $V_L$ -coding vectors are randomly combined via the first restriction site the resulting vectors containing both  $V_H$  and  $V_L$  and both selectable markers can be selected.

Typically the polylinker is a nucleotide sequence that defines one or more, preferably at least two, restriction sites, each unique to the vector, i.e., if it is on the first vector, it is not on the second vector.

**SUBSTITUTE SHEET**

The polylinker restriction sites are oriented to permit ligation of  $V_H$ - or  $V_L$ -coding DNA homologs into the vector in same reading frame as any leader, tag or stop codon sequence present.

5        Random combination is accomplished by ligating  $V_H$ -coding DNA homologs into the first vector, typically at a restriction site or sites within the polylinker. Similarly,  $V_L$ -coding DNA homologs are ligated into the second vector, thereby creating two diverse populations of  
10 expression vectors. It does not matter which type of DNA homolog, i.e.,  $V_H$  or  $V_L$ , is ligated to which vector, but it is preferred, for example, that all  $V_H$ -coding DNA homologs are ligated to either the first or second vector, and all of the  $V_L$ -coding DNA homologs are ligated to the other of  
15 the first or second vector. The members of both populations are then cleaved with an endonuclease at the shared restriction site, typically by digesting both populations with same enzyme. The resulting product is two diverse populations of restriction fragments where the members of  
20 one have cohesive termini complementary to the cohesive termini of the members of the other. The restriction fragments of the two populations are randomly ligated to one another, i.e., a random, interpopulation ligation is performed, to produce a diverse population of vectors each  
25 having a  $V_H$ -coding and  $V_L$ -coding DNA homolog located in the same reading frame and under the control of second vector's promoter. Of course, subsequent recombinations can be effected through cleavage at the shared restriction site, which is typically reformed upon ligation of members  
30 from the two populations, followed by subsequent religations.

The resulting construct is then introduced into an appropriate host to provide amplification and/or expression of the  $V_H$ - and/or  $V_L$ -coding DNA homologs, either  
35 separately or in combination. When coexpressed within the same organism, either on the same or the difference vectors, a functionally active Fv is produced. When the

**SUBSTITUTE SHEET**

$V_H$  and  $V_L$  polypeptides are expressed in different organisms, the respective polypeptides are isolated and then combined in an appropriate medium to form a Fv. Cellular hosts into which a  $V_H$ - and/or  $V_L$ -coding DNA homolog-  
5 containing construct has been introduced are referred to herein as having been "transformed" or as "transformants".

The host cell can be either procaryotic or eucaryotic. Bacterial cells are preferred procaryotic host cells and typically are a strain of E. coli such as, for  
10 example, the E. coli strain DH5 available from Bethesda Research Laboratories, Inc., Bethesda, MD. Preferred eucaryotic host cells include yeast and mammalian cells, preferably vertebrate cells such as those from a mouse, rat, monkey or human cell line.

15 Transformation of appropriate cell hosts with a recombinant DNA molecule of the present invention is accomplished by methods that typically depend on the type of vector used. With regard to transformation of procaryotic host cells, see, for example, Cohen et al.,  
20 Proceedings National Academy of Science, USA Vol. 69, P. 2110 (1972); and Maniatis et al., Molecular Cloning, a Laboratory manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY (1982). With regard to the transformation of vertebrate cells with retroviral vectors  
25 containing rDNAs, see for example, Sorge et al., Mol. Cell. Biol., 4:1730-1737 (1984); Graham et al., Virology, 52:456 (1973); and Wigler et al., Proceedings National Academy of Sciences, USA, Vol. 76, P. 1373-1376 (1979).

b. Expressing the Dicistronic DNA Molecules

30 The dicistronic DNA molecules produced by the above-described method can be operatively linked to a vector for amplification and/or expression.

A variety of methods have been developed to operatively link DNA to vectors via complementary cohesive  
35 termini. For instance, complementary cohesive termini can be engineered into the dicistronic DNA molecules during

the primer extension reaction by use of an appropriately designed polynucleotide synthesis primer, as previously discussed. The dicistronic DNA molecule, and vector if necessary, is cleaved with a restriction endonuclease to produce termini complementary to those of the vector. The complementary cohesive termini of the vector and the dicistronic DNA molecule are then operatively linked (ligated) to produce a unitary double stranded DNA molecule.

10 The present method produces a diverse population of double stranded DNA expression vectors wherein each vector expresses, under the control of a single promoter, one  $V_H$ -coding DNA homolog and one  $V_L$ -coding DNA homolog, the diversity of the population being the result of different  
15  $V_H$ - and  $V_L$ -coding DNA homolog combination that occurs during the PCR reaction where both outside and both inside primers are present in effective amounts. Preferably the vectors are linear double stranded DNA, such as a Lambda Zap derived vector as described herein.

20 In preferred embodiments, the vector defines a nucleotide sequence coding for a ribosome binding site and a leader, the sequence being located downstream from a promoter and upstream from a sequence coding for a polypeptide leader. In preferred embodiments, the vector  
25 contains a selectable marker such that the presence of a dicistronic DNA molecule of this invention inserted into the vector, can be selected. Typical selectable markers are well known to those skilled in the art. Examples of such markers are antibiotic resistance genes, genetically  
30 selectable markers, mutation suppressors such as amber suppressors and the like. The selectable markers are typically located upstream of the promoter.

The resulting construct is then introduced into an appropriate host to provide amplification and/or expression of the  $V_H$ - and  $V_L$ -coding DNA homologs. When  
35 coexpressed within the same organism, a functionally active heterodimeric receptor, such as an  $F_v$ , is produced.

**SUBSTITUTE SHEET**

Cellular hosts into which a  $V_H$ - and  $V_L$ -coding DNA homolog-containing construct has been introduced are referred to herein as having been "transformed" or as "transformants".

The host cell can be either prokaryotic or  
5 eukaryotic. Bacterial cells are preferred prokaryotic host cells for library screening, and typically are a strain of E. coli such as, for example, the E. coli strain DH5 available from Bethesda Research Laboratories, Inc., Bethesda, MD. Preferred eukaryotic host cells include  
10 yeast and mammalian cells, preferably vertebrate cells such as those from a mouse, rat, monkey or human cell line.

Transformation of appropriate cell hosts with a recombinant DNA molecule of the present invention is  
15 accomplished by methods that typically depend on the type of vector used. With regard to transformation of prokaryotic host cells, see, for example, Cohen et al., Proc. Natl. Acad. Sci., USA, 69:2110 (1972); and Maniatis et al., Molecular Cloning: A Laboratory Manual, Cold Spring  
20 Harbor, NY (1982). With regard to the transformation of vertebrate cells with retroviral vectors containing rDNAs, see for example, Sorge et al., Mol. Cell. Biol., 4:1730-1737 (1984); Graham et al., Virology, 52:456 (1973); and Wigler et al., Proc. Natl. Acad. Sci., USA, 76:1373-1376  
25 (1979).

#### 5. Screening For Expression of $V_H$ and/or $V_L$ Polypeptides

Successfully transformed cells, i.e., cells containing a  $V_H$ - and/or  $V_L$ -coding DNA homolog or a dicistronic DNA molecule operatively linked to a vector, can be identified  
30 by any suitable well known technique for detecting the binding of a receptor to a ligand or the presence of a polynucleotide coding for the receptor, preferably its active site. Preferred screening assays are those where the binding of ligand by the receptor produces a detect-  
35 able signal, either directly or indirectly. Such signals include, for example, the production of a complex,

formation of a catalytic reaction product, the release or uptake of energy, and the like. For example, cells from a population subjected to transformation with a subject rDNA can be cloned to produce monoclonal colonies. Cells  
5 form those colonies can be harvested, lysed and their DNA content examined for the presence of the rDNA using a method such as that described Southern, J. Mol. Biol., 98:503 (1975) or Berent et al., Biotech. 3:208 (1985).

In addition to directly assaying for the presence of  
10 a  $V_H$ - and/or  $V_L$ -coding DNA homolog or a dicistronic DNA molecule, successful transformation can be confirmed by well known immunological methods, especially when the  $V_H$  and/or  $V_L$  polypeptides produced contain a preselected epitope. For example, samples of cells suspected of being  
15 transformed are assayed for the presence of the preselected epitope using an antibody against the epitope.

#### 6. $V_H$ - And/Or $V_L$ -Coding Gene Libraries

According to one aspect, the present invention contemplates a gene library, preferably produced by a  
20 primer extension reaction or combination of primer extension reactions as described herein, containing at least about  $10^3$ , preferably at least about  $10^4$  and more preferably at least about  $10^5$  different  $V_H$ - and/or  $V_L$ -coding DNA homologs. The homologs are preferably in an  
25 isolated form, that is, substantially free of materials such as, for example, primer extension reaction agents and/or substrates, genomic DNA segments, and the like.

In preferred embodiments, a substantial portion of the homologs present in the library are operatively linked  
30 to a vector, preferably operatively linked for expression to an expression vector.

Preferably, the homologs are present in a medium suitable for in vitro manipulation, such as water, water containing buffering salts, and the like. The medium  
35 should be compatible with maintaining the activity of the homologs. In addition, the homologs should be present at

a concentration sufficient to allow transformation of a host cell compatible therewith at reasonable frequencies.

It is further preferred that the homologs be present in compatible host cells transformed therewith.

5 C. Expression Vectors

The present invention also contemplates various expression vectors useful in performing, inter alia, the methods of the present invention. Each of the expression vectors is a novel derivative of Lambda Zap vector.

10 1. Lambda Zap II

Lambda Zap II is prepared by replacing the Lambda S gene of the vector Lambda Zap with the Lambda S gene from the Lambda gt10 vector, as described in Example 6.

2. Lambda Zap II  $V_H$

15 Lambda Zap II  $V_H$  is prepared by inserting the synthetic DNA sequences illustrated in Figure 6A into the above-described Lambda Zap II vector. The inserted nucleotide sequence advantageously provides a ribosome binding site (Shine-Dalgarno sequence) to permit proper  
20 initiation of mRNA translation into protein, and a leader sequence to efficiently direct the translated protein to the periplasm. The preparation of Lambda Zap II  $V_H$  is described in more detail in Example 9, and its features illustrated in Figures 6A and 7.

25 3. Lambda Zap II  $V_L$

Lambda Zap II  $V_L$  is prepared as described in Example 12 by inserting into Lambda Zap II the synthetic DNA sequence illustrated in Figure 6B. Important features of Lambda Zap II  $V_L$  are illustrated in Figure 8.



4. Lambda Zap II V<sub>L</sub> II

Lambda Zap II V<sub>L</sub> II is prepared as described in Example 11 by inserting into Lambda Zap II the synthetic DNA sequence illustrated in Figure 10.

5. HCFLP

HCFLP is prepared as described in Example 20 by inserting a flp sequence containing EcoRI compatible ends into the EcoRI site of the lambda Zap II V<sub>H</sub> vector.

6. LCFLP

LCFLP is prepared as described in Example 20 by inserting a flp sequence containing EcoRI compatible ends into the EcoRI site of the lambda Zap II V<sub>L</sub> vector.

7. Lambda ImmunoZAP H

Lambda ImmunoZAP H is prepared by inserting the synthetic DNA sequences illustrated in Figure 25A into the above-described Lambda Zap II vector. The inserted nucleotide sequence advantageously provides a ribosome binding site (Shine-Dalgarno sequence) to permit proper initiation of mRNA translation into protein, and a leader sequence to efficiently direct the translated protein to the periplasm. The preparation of Lambda ImmunoZAP H is described in more detail in Example 28, and its features illustrated in Figures [25A] and [26].

8. Modified Lambda ImmunoZAP H

Modified Lambda ImmunoZAP H is prepared by inserting the modified synthetic DNA sequences illustrated in Figure 8A into the above-described Lambda ZAP II vector. The preparation of modified Lambda ImmunoZAP H and the details of the modifications are described in Example 28B. Its features are illustrated in Figure [24A] and [24B].

### 9. Lambda ImmunoZAP L

Lambda ImmunoZAP L is prepared as described in Example 29 by inserting into Lambda ZAP II the synthetic DNA sequence illustrated in Figure 6B. Important features of Lambda ImmunoZAP L are illustrated in Figure 27.

The above-described vectors are compatible with E. coli hosts, i.e., they can express for secretion into the periplasm proteins coded for by genes to which they have been operatively linked for expression.

### 10 Examples

The following examples are intended to illustrate, but not limit, the scope of the invention.

#### 1. Phenotype Creation

In order to obtain lambda phage clones with a range of desired phenotypes, a combinatorial library selection system was used to generate a diverse collection of clones. This approach utilized two starting populations of lambda phage clones which can be restriction digested, mixed, ligated, and packaged to form a library of clones containing DNA sequences from each of the two populations of parent phage. The following example outlines the method for rapid construction and selection of lambda phage clones containing properties from each of the two parent phage populations derived from lambda WT (cI857 ind1, Sam7) and lambda gt11 (Sam100).

Forty micrograms of a population of lambda phage derived from wild type lambda (WT) DNA (cI857 Sam7) (available from New England Biolabs) was partially digested with lambda HindIII as determined by ethidium bromide staining on 0.8% agarose gels (Maniatis et al., "Molecular Cloning," Cold Spring Harbor Laboratory (1982)). Forty micrograms of a second phage population derived from lambda gt11 DNA (available from Stratagene Cloning Systems, San Diego, CA) was digested to completion with HindIII. Subsequently, this gt11 DNA was digested

**SUBSTITUTE SHEET**

with a second enzyme BamHI in order to reduce the cloning efficiency of the left arm of the gt11 phage (Maniatis et al., supra). Both phage populations had been amplified lytically, which allowed for a relatively high degree of mutations in the resulting DNA. One microgram of the lambda WT DNA was ligated at the Hind III site to 1 to 4  $\mu$ g of the lambda gt11 DNA using T4 DNA ligase in a volume of less than 20  $\mu$ l, according to Maniatis, et al., supra. The ligation mix was subsequently packaged in lambda phage packaging extract, Gigapack<sup>TM</sup> (Stratagene Cloning Systems, San Diego, CA), as described by the manufacturer.

The packaged phage library contained a mixture of many lambda phage constructions. In order to select for desired constructions, phenotypic selection was used to identify those members of the library displaying vigorous growth on supE bacterial hosts. As described by Maniatis et al., supra, dilutions of the phage library was plated with E. coli C600 cells (Stratagene Cloning Systems, San Diego, CA) to generate a lawn of E. coli with isolated lambda plaques. These isolated plaques are result of clonal expansion from a single lambda phage clone. Since C600 cells are supE, the growth vigor of the individual lambda phage clones could be assessed by the size of the lambda phage plaque on the E. coli lawn. The parental WT phage do not form plaques on E. coli C600. At least three classes of phage were identified and subsequently categorized as small, medium, or large plaque size. The large plaque size was an indication of vigorous growth on the phage lawn, while small plaque size indicated poor growth. This demonstrates selection for the phenotype of the S gene based on plaque size. Other phenotypes could be used for selection.

Subsequent characterization by restriction mapping and plating on sup0 (these strains contain no amber codon suppressing tRNAs) and supF E. coli hosts, indicated that at least one of the large plaque forming clones, L2, did not contain an amber mutation as found in the lambda WT

(Sam7) or lambda gt11 (C5100) parent phage. One of the small plaque phage, S2, contained the left arm of lambda WT gene and the right arm of lambda gt11 containing the Sam100 gene. This Sam100 mutation is known to grow poorly  
5 on supE hosts and is optimal on a supF strain, with no growth on a supO host. The remaining library of clones displayed several different phenotypes, dictated by the diversity of the two starting populations of phage. Some clones also exhibited phenotypes that resulted from the  
10 random assortment of two mutant DNA fragments derived from just one of the parent DNA molecules. This illustrates the concept that the two genes that give rise to the populations of interest need not be on separate DNA molecules at the start of the method.

15 Due to the phenotypic selection applied following the ligation and packaging of the phage library, the large diversity of these two populations of phage was not completely analyzed. However, the range of clones identified with alternate S gene phenotypes demonstrated some of this  
20 diversity. The diversity in these two populations of lambda phage is believed to be derived from the low level of spontaneous mutations which occur through repeated rounds of replication required in large scale preparations of lambda phage. However, the spontaneous mutations  
25 occurring within each of these individual phage populations could not generate a collection of lambda phage containing characteristics of both parent populations of phage. This combinatorial approach, therefore, provides a mechanism in which novel constructions can be generated  
30 that express genes from both parent phage constructions.

## 2. Polynucleotide Selection for Immunoglobulin Production

The nucleotide sequences encoding the immunoglobulin protein CDR's are highly variable. However, there are several regions of conserved sequences that flank the V<sub>H</sub>  
35 domains. For instance, contain substantially conserved nucleotide sequences, i.e., sequences that will hybridize

to the same primer sequence. Therefore, polynucleotide synthesis (amplification) primers that hybridize to the conserved sequences and incorporate restriction sites into the DNA homolog produced that are suitable for operatively linking the synthesized DNA fragments to a vector were constructed. More specifically, the DNA homologs were inserted into lambda Zap II vector (Stratagene Cloning System, San Diego, CA) at the XhoI and EcoRI sites. For amplification of the  $V_H$  domains, the 3' primer (primer 67 in Table 7), was designed to be complementary to the mRNA in the  $J_H$  region. In all cases, the 5' primers (primers 56-65,, Table 7) were chosen to be complementary to the first strand cDNA in the conserved N-terminus region (antisense strand). Initially amplification was performed with a mixture of 32 primers (primer 56, Table 7) that were degenerate at five positions. Hybridoma mRNA could be amplified with mixed primers, but initial attempts to amplify mRNA from spleen yielded variable results. Therefore, several alternatives to amplification using the mixed 5' primers were compared.

The first alternative was to construct multiple unique primers, eight of which are shown in Table 7, corresponding to individual members of the mixed primer pool. The individual primers 52-64 of Table 7 were constructed by incorporating either of the two possible nucleotides at three of the five degenerate positions.

The second alternative was to construct a primer containing inosine (primer 65, table 7) at four of the variable positions based on the published work of Takahashi, et al., Proc. Natl. Acad. Sci. (U.S.A.), 82:1931-1935, (1985) and Ohtsuka et al., J. Biol. Chem., 260:2605-2608, (1985). This primer has the advantage that it is not degenerate and, at the same time minimizes the negative effects of mismatches at the unconserved positions as discussed by Martin et al., Nu. Acids Res., 13:8927 (1985). However, it was not known if the presence of inosine nucleotides would result in incorporation of

**SUBSTITUTE SHEET**

unwanted sequences in the cloned  $V_H$  regions. Therefore, inosine was not included at the one position that remains in the amplified fragments after the cleavage of the restriction sites. As a result, inosine was not in the  
5 cloned insert.

Additional,  $V_H$  amplification primers including the unique 3' primer were designed to be complementary to a portion of the first constant region domain of the gamma 1 heavy chain mRNA (Primers 70 and 71, Table 7). These  
10 primers will produce DNA homologs containing polynucleotides coding for amino acids from the  $V_H$  and the first constant region domains of the heavy chain. These DNA homologs can therefore be used to produce Fab fragments rather than an  $F_v$ .

15 As a control for amplification from spleen or hybridoma mRNA, a set of primers hybridizing to a highly conserved region within the constant region IgG, heavy chain gene were constructed. The 5' primer (primer 66, Table 7) is complementary to the cDNA in the  $C_H2$  region  
20 whereas the 3' primer (primer 68, Table 7) is complementary to the mRNA in the  $C_H3$  region. It is believed that no mismatches were present between these primers and their templates.

The nucleotide sequences encoding the  $V_L$  CDRs are  
25 highly variable. However, there are several regions of conserved sequences that flank the  $V_L$  CDR domains including the  $J_L$ ,  $V_L$  framework regions and  $V_L$  leader/promoter. Therefore, amplification primers that hybridize to the conserved sequences and incorporate restriction sites that  
30 allowing cloning the amplified fragments into the pBluescript SK-vector cut with Nco I and SpeI were constructed. For amplification of the  $V_L$  CDR domains, the 3' primer (primer 69 in Table 7), was designed to be complementary to the mRNA in the  $J_L$  regions. The 5' primer  
35 (primer 70, Table 7) was chosen to be complementary to the first strand cDNA in the conserved N-terminus region (antisense strand).

A second set of amplification primers for amplification of the  $V_L$  CDR domains the 5' primers (primers 73-80 in Table 8) were designed to be complementary to the first strand cDNA in the conserved N-terminus region. These  
5 primers also introduced a Sac I restriction endonuclease site to allow the FLDNA homolog to be cloned into the  $V_{LII}$ -expression vector. The 3'  $V_L$  amplification primer (primer 81 in Table 8) was designed to be complementary to the mRNA in the  $J_L$  regions and to introduce the XbaI restriction  
10 endonuclease site required to insert the  $V_L$  DNA homolog into the  $V_{LII}$ -expression vector (Figure 8).

Additional 3'  $V_L$  amplification primers were designed to hybridize to the constant region of either kappa or lambda mRNA (primers 82 and 83 in Table 8). These primers  
15 allow a DNA homolog to be produced containing polynucleotide sequences coding for constant region amino acids of either kappa or lambda chain. These primers make it possible to produce an Fab fragment rather than an  $F_v$ .

The primers used for amplification of kappa light  
20 chain sequences for construction of Fabs are shown at least in Table 8. Amplification with these primers was performed in 5 separate reactions, each containing one of the 5' primers (primers 75-78, and 84) and one of the 3' primer (primer 81) has been used to construct  $F_v$  fragments.  
25 The 5' primers contain a Sac I restriction site and the 3' primers contain a XbaI restriction site.

The primers used for amplification of heavy chain Fd fragments for construction of Fabs are shown at least in Table 7. Amplification was performed in eight separate  
30 reactions, each containing one of the 5' primers (primers 57-64) and one of the 3' primers (primer 70). The remaining 5' primers that have been used for amplification in a single reaction are either a degenerate primer (primer 56) or a primer that incorporates inosine at four degenerate  
35 positions (primer 66, Table 7, and primers 89 and 90, Table 8). The remaining 3' primer (primer 86, Table 8) has been used to construct  $F_v$  fragments. Many of the 5'

primers incorporate a Xho I site, and 3' primers include a SpeI restriction site.

V<sub>L</sub> amplification primers designed to amplify human light chain variable regions of both the lambda and kappa  
5 isotypes are also shown in Table 8.

All primers and synthetic polynucleotides used herein and shown on Tables 7-11 were either purchased from Research Genetics in Huntsville, Alabama or synthesized on  
an Applied Biosystems DNA synthesizer, model 381A, using  
10 the manufacturer's instructions.

**SUBSTITUTE SHEET**



TABLE 7

5	(56) 5' AGGT(C/G)(C/A)A(G/A)CT(G/T)CTCGAGTC(T/A)GG 3'	degenerate 5' primer for the amplification of variable heavy chain region (V <sub>H</sub> )
	(57) 5' AGGTCCAGCTGCTCGAGTCTGG 3'	Unique 5' primer for the amplification of V <sub>H</sub>
	(58) 5' AGGTCCAGCTGCTCGAGTCAGG 3'	"
	(59) 5' AGGTCCAGCTTCTCGAGTCTGG 3'	"
	(60) 5' AGGTCCAGCTTCTCGAGTCAGG 3'	"
10	(61) 5' AGGTCCAACCTGCTCGAGTCTGG 3'	"
	(62) 5' AGGTCCAACCTGCTCGAGTCAGG 3'	"
	(63) 5' AGGTCCAACCTTCTCGAGTCTGG 3'	"
	(64) 5' AGGTCCAACCTTCTCGAGTCAGG 3'	"
15	(65) 5' AGGTIIAICTTCTCGAGTC(T/A) 3'	5' degenerate primer containing inosine at 4 degenerate positions
	(66) 5' GCCCAAGGATGTGCTCACC 3'	5' primer for amplification in the C <sub>H</sub> 2 region of mouse IgG1
	(67) 5' CTATTAGAAATTCACGGTAACAGTGGTGCCTTGGCCCCCA 3'	3' primer for amplification of V <sub>H</sub>
	(67A) 5' CTATTAACTAGTAACGGTAACAGTGGTGCCTTGGCCCCCA 3'	3' primer for amplification of V <sub>H</sub> using 3' spe I site
20	(68) 5' CTCAGTATGGTGGTTGTGC 3'	3' primer for amplification in the C <sub>H</sub> 3 region of mouse IgG1
	(69) 5' GCTACTAGTTTGTGATTTCCACCTTGG 3'	3' primer for amplification of V <sub>L</sub>

- (70) 5' CAGCCATGGCCGACATCCAGATG 3' 5' primer for amplification of  $V_L$   
 (71) 5' AATTCTACTAGTCACCTTGGTGCTGCTGGC 3' Unique 3' primer for amplification of  $V_H$  including part of the mouse gamma 1 first constant  
 5 (72) 5' TATGCAACTAGTACAACCAATCCCTGGGCACAATTTT 3' Unique 3' primer for amplification of  $V_H$  including part of mouse gamma 1 first constant region and hinge region

TABLE 8

	(73)	5'	CCAGTTCGGAGCTCGTTGTGACTCAGGAATCT	3'	Unique 5' primer for the amplification of $V_L$
	(74)	5'	CCAGTTCGGAGCTCGTTGTGACGCGCGCCC	3'	"
5	(75)	5'	CCAGTTCGGAGCTCGTGTCTCAGCCAGTCTCCA	3'	"
	(76)	5'	CCAGTTCGGAGCTCCAGATGACCCAGTCTCCA	3'	"
	(77)	5'	CCAGATGTGAGCTCGTGATGACCCAGACTCCA	3'	"
	(78)	5'	CCAGATGTGAGCTCGTCAATGACCCAGTCTCCA	3'	"
	(79)	5'	CCAGATGTGAGCTCTTGATGACCCAACTCAA	3'	"
10	(80)	5'	CCAGATGTGAGCTCGTGATAACCCAGGATGAA	3'	"
	(81)	5'	GCAGCATCTAGAGTTTCAGCTCCAGCTTGCC	3'	Unique 3' primer for $V_L$ amplification <sup>8</sup>
	(82)	5'	CCGCCGCTAGAACACTCATTCCTGTGAAGCT	3'	Unique 3' primer for $V_L$ amplification including the kappa constant region
15	(83)	5'	CCGCCGCTAGAACATTCTGCAGGAGACAGACT	3'	Unique 3' primer for $V_L$ amplification including the lambda constant region
	(84)	5'	CCAGTTCGGAGCTCGTGATGACACAGTCTCCA	3'	Unique 5' primer for $V_L$ amplification
	(85)	5'	GCGCCGCTAGAAATTAACACTCATTCCTGTGAA	3'	Unique 3' primer for $V_L$ amplification
	(86)	5'	CTATTAACTAGTAACGGTAACAGTGGTGCCTTGCCCCA	3'	"
	(87)	5'	AGGCTTACTAGTACAAATCCCTGGGCACAAT	3'	Unique 3' primer for $V_H$ amplification
20	(88)	5'	GCGGCTCTAGAACACTCATTCCTGTGAA	3'	Unique 3' primer for $V_L$ amplification
	(89)	5'	AGGTIIAICTICTCGAGTCTGC	3'	Degenerate 5' primer containing inosine at 4 degenerate positions
	(90)	5'	AGGTIIAICTICTCGAGTCAGC	3'	"

3. Production Of A  $V_H$  Coding Repertoire Enriched In  
FITC Binding Proteins

Fluorescein isothiocyanate (FITC) was selected as a ligand for receptor binding. It was further decided to  
5 enrich by immunization the immunological gene repertoire, i.e.,  $V_H$ - and  $V_L$ -coding gene repertoires, for genes coding for anti-FITC receptors. This was accomplished by linking FITC to keyhole limpet hemocyanin (KLH) using the techniques described in Antibodies A Laboratory Manual, Harlow  
10 and Lowe, eds., Cold Spring Harbor, New York, (1988). Briefly, 10.0 milligrams (mg) of keyhole limpet hemocyanin and 0.5 mg of FITC were added to 1 ml of buffer containing 0.1 M sodium carbonate at pH 9.6 and stirred for 18 to 24 hours at 4 degrees C (4C). The unbound FITC was removed  
15 by gel filtration through Sephadex G-25.

The KLH-FITC conjugate was prepared for injunction into mice by adding 100  $\mu$ g of the conjugate to 250  $\mu$ l of phosphate buffered saline (PBS). An equal volume of complete Freund's adjuvant was added and the entire solution  
20 was emulsified for 5 minutes. A 129  $G_{1X+}$  mouse was injected with 300  $\mu$ l of the emulsion. Injections were given subcutaneously at several sites using a 21 gauge needle. A second immunization with KLH-FITC was given two week later. This injection was prepared as follows:  
25 fifty  $\mu$ g of KLH-FITC were diluted in 250  $\mu$ L of PBS and an equal volume of alum was admixed to the KLH-FITC solution. The mouse was injected intraperitoneally with 500  $\mu$ l of the solution using a 23 gauge needle. One month later the mice were given a final injection of 50  $\mu$ g of the KLH-  
30 FITC conjugate diluted to 200  $\mu$ L in PBS. This injection was given intravenously in the lateral tail vein using a 30 gauge needle. Five days after this final injection the mice were sacrificed and total cellular RNA was isolated from their spleens.

35 Hybridoma PCP 8D11 producing an antibody immuno-specific for phosphonate ester was cultured in DMEM media (Gibco Laboratories, Grand Island, New York) containing 10

percent fetal calf serum supplemented with penicillin and streptomycin. About  $5 \times 10^8$  hybridoma cells were harvested and washed twice in phosphate buffered saline. Total cellular RNA was prepared from these isolated hybridoma  
5 cells.

#### 4. Preparation Of A $V_H$ -Coding Gene Repertoire

Total cellular RNA was prepared from the spleen of a single mouse immunized with KLH-FITC as described in Example 3 using the RNA preparation methods described by  
10 Chomczynski et al., Anal Biochem., 162:156-159 (1987) using the manufacturer's instructions and the RNA isolation kit produced by Stratagene Cloning Systems, La Jolla, CA. Briefly, immediately after removing the spleen from the immunized mouse, the tissue was homogenized in 10 ml  
15 of a denaturing solution containing 4.0 M guanine isothiocyanate, 0.25 M sodium citrate at pH 7.0, and 0.1 M 2-mercaptoethanol using a glass homogenizer. One ml of sodium acetate at a concentration of 2 M at pH 4.0 was admixed with the homogenized spleen. One ml of phenol  
20 that had been previously saturated with  $H_2O$  was also admixed to the denaturing solution containing the homogenized spleen. Two ml of a chloroform:isoamyl alcohol (24:1 v/v) mixture was added to this homogenate. The homogenate was mixed vigorously for ten seconds and  
25 maintained on ice for 15 minutes. The homogenate was then transferred to a thick-walled 50 ml polypropylene centrifuge tube (Fisher Scientific Company, Pittsburgh, PA). The solution was centrifuged at  $10,000 \times g$  for 20 minutes at  $4^\circ C$ . The upper RNA-containing aqueous layer was  
30 transferred to a fresh 50 ml polypropylene centrifuge tube and mixed with an equal volume of isopropyl alcohol. This solution was maintained at  $-20^\circ C$  for at least one hour to precipitate the RNA. The solution containing the precipitated RNA was centrifuged at  $10,000 \times g$  for twenty minutes  
35 at  $4^\circ C$ . The pelleted total cellular RNA was collected and dissolved in 3 ml of the denaturing solution described

**SUBSTITUTE SHEET**

above. Three ml of isopropyl alcohol was added to the resuspended total cellular RNA and vigorously mixed. This solution was maintained at  $-20^{\circ}\text{C}$  for at least 1 hour to precipitate the RNA. The solution containing the precipitated RNA was centrifuged at  $10,000 \times g$  for ten minutes at  $4^{\circ}\text{C}$ . The pelleted RNA was washed once with a solution containing 75% ethanol. The pelleted RNA was dried under vacuum for 15 minutes and then resuspended to dimethyl pyrocarbonate (DEPC) treated  $\text{H}_2\text{O}$  (DEPC- $\text{H}_2\text{O}$ ).

10 Messenger RNA (mRNA) enriched for sequences containing long poly A tracts was prepared from the total cellular RNA using methods described in Molecular Cloning A Laboratory Manual, Maniatis et al., eds. Cold Spring Harbor Laboratory, New York, (1982). Briefly, one half of  
15 the total RNA isolated from a single immunized mouse spleen prepared as described above was resuspended in one ml of DEPC- $\text{H}_2\text{O}$  and maintained at  $65^{\circ}\text{C}$  for five minutes. One ml of 2x high salt loading buffer consisting of 100 mM Tris-HCl, 1M sodium chloride, 2.0 mM disodium ethylene  
20 diamine tetraacetic acid (EDTA) at pH 7.5, and 0.2% sodium dodecyl sulfate (SDS) was added to the resuspended RNA and the mixture allowed to cool to room temperature. The mixture was then applied to an oligo-dT (Collaborative Research Type 2 or Type 3) column that was previously  
25 prepared by washing the oligo-dT with a solution containing 0.1 M sodium hydroxide and 5 mM EDTA and then equilibrating the column with DEPC- $\text{H}_2\text{O}$ . The eluate was collected in a sterile polypropylene tube and reapplied to the same column after heating the eluate for 5 minutes at  
30  $65^{\circ}\text{C}$ . The oligo dT column was then washed with 2 ml of high salt loading buffer consisting of 50 mM Tris-HCl at pH 7.5, 500 mM sodium chloride, 1 mM EDTA at pH 7.5 and 0.1% SDS. The oligo dT column was then washed with 2 ml of 1 X medium salt buffer consisting of 50 mM Tris-HCl at  
35 pH 7.5, 100 mM sodium chloride 1 mM EDTA and 0.1% SDS. The messenger RNA was eluted from the oligo dT column with 1ml of buffer consisting of 10 mM Tris-HCL at pH 7.5, 1 mM

**SUBSTITUTE SHEET**

EDTA at pH 7.5 and 0.05% SDS. The messenger RNA was purified by extracting this solution with phenol/chloroform followed by a single extraction with 100% chloroform. The messenger RNA was concentrated by ethanol precipitation and resuspended in DEPC- H<sub>2</sub>O.

The messenger RNA isolated by the above process contains a plurality of different V<sub>H</sub> coding polynucleotides, i.e., greater than about 10<sup>4</sup> different V<sub>H</sub>-coding genes.

10 5. Preparation Of A Single V<sub>H</sub> Coding Polynucleotide

Polynucleotides coding for a single V<sub>H</sub> were isolated according to Example 4 except total cellular RNA was extracted from monoclonal hybridoma cells prepared in Example 3. The polynucleotides isolated in this manner code for a single V<sub>H</sub>.

6. DNA Homolog Preparation

In preparation for PCR amplification, mRNA prepared according to the above examples was used as a template for cDNA synthesis by a primer extension reaction. In a typical 50 ul transcription reaction, 5-10 ug of spleen or hybridoma mRNA in water was first hybridized (annealed) with 500 ng (50.0 pmol) of the 3' V<sub>H</sub> primer (primer 67, Table 7), at 65°C for five minutes. Subsequently, the mixture was adjusted to 1.5 mM dATP, dCTP, dGTP and dTTP, 40 mM Tris-HCl at pH 8.0, 8 mM MgCl<sub>2</sub>, 50 mM NaCl, and 2 mM spermidine. Moloney-Murine Leukemia virus Reverse transcriptase (Stratagene Cloning Systems), 26 units, was added and the solution was maintained for 1 hours at 37°C.

PCR amplification was performed in a 100 ul reaction containing the products of the reverse transcription reaction (approximately 5 ug of the cDNA/RNA hybrid), 300 ng of 3' V<sub>H</sub> primer (primer 67 of Table 7), 300 ng each of the 5' V<sub>H</sub> primers (primer 57-65 of Table 7) 200 mM of a mixture of dNTP's, 50 mM KCl, 10 mM Tris-HCl pH 8.3, 15 mM MgCl<sub>2</sub>, 0.1% gelatin and 2 units of Taq DNA polymerase. The

**SUBSTITUTE SHEET**

reaction mixture was overlaid with mineral oil and subjected to 40 cycles of amplification. Each amplification cycle involved denaturation at 92°C for 1 minute, annealing at 52°C for 2 minutes and polynucleotide synthesis by  
5 Primer extension (elongation) at 72°C for 1.5 minutes. The amplified  $V_H$ -coding DNA homolog containing samples were extracted twice with phenol/chloroform, once with chloroform, ethanol precipitated and were stored at -70°C in 10 mM Tris-HCl, (pH, 7.5) and 1 mM EDTA.

10 Using unique 5' primers (57-64, Table 7), efficient  $V_H$ -coding DNA homolog synthesis and amplification from the spleen mRNA was achieved as shown in Figure 3, lanes R17-R24. The amplified cDNA ( $V_H$ -coding DNA homolog) is seen as a major band of the expected size (360 bp). The intensi-  
15 ties of the amplified  $V_H$ -coding polynucleotide fragment in each reaction appear to be similar, indicating that all of these primers are about equally efficient in initiating amplification. The yield and quality of the amplification with these primers was reproducible.

20 The primer containing inosine also synthesized amplified  $V_H$ -coding DNA homologs from spleen mRNA reproducibly, leading to the production of the expected sized fragment, of an intensity similar to that of the other amplified cDNAs (Figure 4, lane R16). This result indicated that  
25 the presence of inosine also permits efficient DNA homolog synthesis and amplification. Clearly indicating how useful such primers are in generating a plurality of  $V_H$ -coding DNA homologs. Amplification products obtained from the constant region primers (primers 66 and 68, Table 7)  
30 were more intense indicating that amplification was more efficient, possibly because of a higher degree of homology between the template and primers (Figure 4, Lang R9). Based on these results, a  $V_H$ -coding gene library was constructed from the products of eight amplifications,  
35 each performed with a different 5' primer. Equal portions of the products from each primer extension reaction were

**SUBSTITUTE SHEET**



mixed and the mixed product was then used to generate a library of  $V_H$ -coding DNA homolog-containing vectors.

DNA homologs of the  $V_L$  were prepared from the purified mRNA prepared as described above. In preparation for PCR amplification, mRNA prepared according to the above examples was used as a template for cDNA synthesis. In a typical 50 ul transcription reaction, 5-10 ug of spleen or hybridoma mRNA in water was first annealed with 300 ng (50.0 pmol) of the 3'  $V_L$  primer (primer 69, Table 7), at 65°C for five minutes. Subsequently, the mixture was adjusted to 1.5 mM dATP, dCTP, dGTP, and dTTP, 40 mM Tris-HCL at pH 8.0, 8 mM  $MgCl_2$ , 50 mM NaCl, and 2 mM spermidine. Moloney-Murine Leukemia virus reverse transcriptase (Stratagene Cloning Systems), 26 units, was added and the solution was maintained for 1 hour at 37°C. The PCR amplification was performed in a 100 ul reaction containing approximately 5 ug of the cDNA/RNA hybrid produced as described above, 300 ng of the 3'  $V_L$  primer (primer 69 of Table 7), 300 ng of the 5'  $V_L$  primer (primer 70 of Table 7), 200 mM of a mixture of dNTP's, 50 mM KCl, 10 mM Tris-HCl pH 8.3, 15 mM  $MgCl_2$ , 0.1% gelatin and 2 units of Taq DNA polymerase. The reaction mixture was overlaid with mineral oil and subjected to 40 cycles of amplification. Each amplification cycle involved denaturation at 92°C for 1 minute, annealing at 52°C for 2 minutes and elongation at 72°C for 1.5 minutes. The amplified samples were extracted twice with phenol/chloroform, once with chloroform, ethanol precipitated and were stored at 70°C in 10 mM Tris-HCl at pH 7.5 and 1 mM EDTA.

#### 7. Inserting DNA Homologs Into Vectors

In preparation for cloning a library enriched in  $V_H$  sequences, PCR amplified products (2.5 mg/30 ul of 150 mM NaCl, 8 mM Tris-HCl (pH 7.5), 6 mM  $MgSO_4$ , 1 mM DTT, 200 mg/ml bovine serum albumin (BSA) at 37°C were digested with restriction enzymes Xho I (125 units) and EcoR I (10 U) and purified on a 1% agarose gel. In cloning experi-

ments which required a mixture of the products of the amplification reactions, equal volumes (50 ul, 1-10 ug concentration) of each reaction mixture were combined after amplification but before restriction digestion.

5 After gel electrophoresis of the digested PCR amplified spleen mRNA, the region of the gel containing DNA fragments of approximately 350 bps was excised, electroeluted into a dialysis membrane, ethanol precipitated and resuspended in 10 mM Tris-HCl pH 7.5 and 1 mM EDTA to a final

10 concentration of 10 ng/ul. Equimolar amounts of the insert were then ligated overnight at 5°C to 1 ug of Lambda Zap<sup>TM</sup> II vector (Stratagene Cloning Systems, La Jolla, CA) previously cut by EcoR I and Xho I. A portion of the ligation mixture (1 ul) was packaged for 2 hours at

15 room temperature using Gigapack Gold packaging extract (Stratagene Cloning Systems, La Jolla, CA), and the packaged material was plated on IL1-blue host cells. The Library was determined to consist of  $2 \times 10^7$  V<sub>H</sub> homologs with less than 30% non-recombinant background.

20 The vector used above, Lambda Zap II is a derivative of the original Lambda Zap (ATCC # 40,298) that maintains all of the characteristics of the original Lambda Zap including 6 unique cloning sites, fusion protein expression, and the ability to rapidly excise the insert in the

25 form of a phagemid (Bluescript SK-), but lacks the SAM 100 mutation, allowing growth on many Non-Sup F strains, including XL1-Blue. The Lambda Zap II was constructed as described in Short et al., Nucleic Acids Res., 16:7583-7600, (1988), by replacing the Lambda S gene contained in

30 a 4254 base pair (bp) DNA fragment produced by digesting Lambda Zap with the restriction enzyme NcoI. This 4254 bp DNA fragment was replaced with the 4254 bp DNA fragment containing the Lambda S gene isolated from Lambda gt10 (ATCC # 40,179) after digesting the vector with the

35 restriction enzyme NcoI. The 4254 bp DNA fragment isolated from lambda gt10 was ligated into the original Lambda Zap vector using T4 DNA ligase and standard proto-

cols for such procedures described in Current Protocols in Molecular Biology, Ausubel et al., eds., John Wiley and Sons, New York, (1987).

In preparation of cloning a library enriched in  $V_L$  sequences, 2 ug of PCR amplified products (2.5 mg/30 ul of 150 mM NaCl, 8 mM Tris-HCl (pH 7.5), 6 mM  $MgSO_4$ , 1 mM DTT, 200 mg/ml BSA) were digested with restriction enzymes Nco I (30 unites) and Spe I (45 units) at 37°C for 2 hours. The digested PCR amplified products were purified on 1% agarose gel using standard electroelution technique described in Molecular Cloning A Laboratory Manual, Maniatis et al., eds., Cold Spring Harbor, New York, (1982). Briefly, after gel electroelution of the digested PCR amplified product the region of the gel containing the  $V_L$ -coding DNA fragment of the appropriate size was excised, electroelution into a dialysis membrane, ethanol precipitated and resuspended at a final concentration of 10 ng per ml in a solution containing 10 mM Tris-HCL at pH 7.5 and 1 mM EDTA.

An equal molar amount of DNA representing a plurality of different  $V_L$ -coding DNA homologs was ligated to a pBluescript SK- phagemid vector that had been previously cut with Nco I and Spe I. A portion of the ligation mixture was transformed using the manufacturer's instructions into Epicurian Coli XL1-Blue competent cells (Stratagene Cloning Systems, La Jolla, CA). The transformant library was determined to consist of  $1.2 \times 10^3$  colony forming units/ug of  $V_L$  homologs with less than 3% non-recombinant background.

#### 8. Sequencing of Plasmids From the $V_L$ -Coding cDNA Library

To analyze the Lambda Zap II phage clones, the clones were excised from Lambda Zap into plasmids according to the manufacture's instructions (Stratagene Cloning System, La Jolla, CA). Briefly, phage plaques were cored from the agar plates and transferred to sterile microfuge tubes containing 500 ul a buffer containing 50 mM Tris-HCL at pH

7.5, 100 mM NaCl, 10 mM MgSO<sub>4</sub>, and 0.01% gelatin and 20 uL of Chloroform.

For excisions, 200 ul of the phage stock, 200 ul of XL1-Blue cells ( $A_{600} = 1.00$ ) and 1 ul of R408 helper phage  
5 (1 x 10<sup>11</sup> pfu/ml) were incubated at 37°C for 15 minutes. The excised plasmids were infected into XL1-Blue cells and plated onto LB plates containing ampicillin. Double stranded DNA was prepared from the phagemid containing cells according to the methods described by Holmes et al.,  
10 Anal. Biochem., 114:193, (1981). Clones were first screened for DNA inserts by restriction digests with either Pvu II or Bgl I and clones containing the putative V<sub>H</sub> insert were sequenced using reverse transcriptase according to the general method described by Sanger et  
15 al., Proc. Natl. Acad. Sci., USA, 74:5463-5467, (1977) and the specific modifications of this method provided in the manufacturer's instruction in the AMV reverse transcriptase <sup>35</sup>S-dATP sequencing kit from Stratagene Cloning Systems, La Jolla, CA.

20 9. Characterization Of The Cloned V<sub>H</sub> Repertoire

The amplified products which had been digested with Xho I and EcoR I and cloned into Lambda Zap, resulted in a cDNA library with 9.0 x 10<sup>5</sup> pfu's. In order to confirm that the library consisted of a diverse population of V<sub>H</sub>-  
25 coding DNA homologs, the N-terminal 120 bases of 18 clones, selected at random from the library, were excised and sequenced (Figure 5). To determine if the clones were of V<sub>H</sub> gene origin, the cloned sequences were compared with known V<sub>H</sub> sequences and V<sub>L</sub> sequences. The clones exhibited  
30 from 80 to 90% homology with sequences of known heavy chain origin and little homology with sequences of light chain origin when compared with the sequences available in Sequences of Proteins of Immunological Interest by Kabot et al., 4th ed., U.S. Dept. of Health and Human Sciences,  
35 (1987). This demonstrated that the library was enriched

for the desired  $V_H$  sequence in preference to other sequences, such as light chain sequences.

The diversity of the population was assessed by classifying the sequenced clones into predefined subgroups (Figure 5). Mouse  $V_H$  sequences are classified into eleven subgroups (Figure 5). Mouse  $V_H$  sequences are classified into eleven subgroups [I (A,B), II (A,B,C), III (A,B,C,D(V (A,B))] based on framework amino acid sequences described in Sequences of proteins of Immunological Interest by Kabot et al., 4th ed., U.S. Dept. of Health and Human Sciences, (1987); Dildrop, Immunology Today, 5:84, (1984); and Brodeur et al., Eur. J. Immunol., 14:922, (1984). Classification of the sequenced clones demonstrated that the cDNA library contained  $V_H$  sequences of at least 7 different subgroups. Further, a pairwise comparison of the homology between the sequenced clones showed that no two sequences were identical at all positions, suggesting that the population is diverse to the extent that it is possible to characterize by sequence analysis.

Six of the clones (L 36-50, Figure 5) belong to the subclass III B and had very similar nucleotide sequences. This may reflect a preponderance of mRNA derived from one or several related variable genes in stimulated spleen, but the data does not permit ruling out the possibility of a bias in the amplification process.

#### 10. $V_H$ -Expression Vector Construction

The main criterion used in choosing a vector system was the necessity of generating the largest number of Fab fragments which could be screened directly. Bacteriophage lambda was selected as the expression vector for three reasons. First, in vitro packaging of phage DNA is a highly efficient method of reintroducing DNA into host cells. Second, it is possible to detect protein expression at the level of single phage plaques. Finally, the screening of phage libraries typically involve less difficulty with nonspecific binding. An alternative,

plasmid cloning vectors, are only advantageous in the analysis of clones after they have been identified. This advantage is not lost in the present system because of the use of lambda Zap, thereby permitting a plasmid containing the heavy chain, light chain, or Fab expressing inserts to be excised.

To express the plurality of  $V_H$ -coding DNA homologs in an E. coli host cell, a vector was constructed that placed the  $V_H$ -coding DNA homologs in the proper reading frame, provided a ribosome binding site as described by Shine et al., Nature, 254:34, 1975, provided a leader sequence directing the expressed protein to the periplasmic space, provided a polynucleotide sequence that coded for a known epitope (epitope tag) and also provided a polynucleotide that coded for a spacer protein between the  $V_H$ -coding DNA homolog and the polynucleotide coding for the epitope tag. A synthetic DNA sequence containing all of the above polynucleotides and features was constructed by designing single stranded polynucleotide segments of 20-40 bases that would hybridize to each other and form the double stranded synthetic DNA sequence shown in Figure 6. The individual single-stranded polynucleotides (N1-N12) are shown in Table 9.

Polynucleotides N2, N3, N9-4', N11, N10-5', N6, N7 and N8 were kinased by adding 1 ul of each polynucleotide (0.1 ug/ul) and 20 units of T4 polynucleotide kinase to a solution containing 70 mM Tris-HCl at pH 7.6, 10 mM  $MgCl_2$ , 5 mM DTT, 10 mM 2-mercaptoethanol (2ME), 500 micrograms per ml of BSA. The solution was maintained at 37°C for 30 minutes and the reaction stopped by maintaining the solution at 65°C for 10 minutes. The two end polynucleotides 20 ng of polynucleotides N1 and polynucleotides N12, were added to the above kinasing reaction solution together with 1/10 volume of a solution containing 20.0 mM Tris-HCl at pH 7.4, 2.0 mM  $MgCl_2$  and 50.0 mM NaCl. This solution was heated to 70°C for 5 minutes and allowed to cool to room temperature, approximately 25°C, over 1.5 hours in

**SUBSTITUTE SHEET**

a 500 ml beaker of water. During this time period all 10 polynucleotides annealed to form the double stranded synthetic DNA insert shown in Figure 6A. The individual polynucleotides were covalently linked to each other to stabilize the synthetic DNA insert by adding 40 ul of the above reaction to a solution containing 50 mM Tris-HCl at pH 7.5, 7 mM MgCl<sub>2</sub>, 1 mM DTT, 1 mM adenosine triphosphate (ATP) and 10 units of T4 DNA ligase. This solution was maintained at 37°C for 30 minutes and then the T4 DNA ligase was inactivated by maintaining the solution at 65°C for 10 minutes. The end polynucleotides were kinased by mixing 52 ul of the above reaction, 4 ul of a solution containing 10 mM ATP and 5 units of T4 polynucleotide kinase. This solution was maintained at 37°C for 30 minutes. The completed synthetic DNA insert was ligated directly into a lambda Zap II vector that had been previously digested with the restriction enzymes Not I and Xho I. The ligation mixture was packaged according to the manufacture's instructions using Gigapack II Gold packing extract available from Stratagene Cloning Systems, La Jolla, CA. The packaged ligation mixture was plated on XL1 blue cells (Stratagene Cloning Systems, San Diego, CA). Individual lambda Zap II plaques were cored and the inserts excised according to the in vivo excision protocol provided by the manufacturer, Stratagene Cloning Systems, La Jolla, CA. This in vivo excision protocol moves the cloned insert from the lambda Zap II vector into a plasmid vector to allow easy manipulation and sequencing. The accuracy of the above cloning steps was confirmed by sequencing the insert using the Sanger dideoxide method described in by Sanger et al., Proc. Natl. Acad. Sci USA, 74:5463-5467, (1977) and using the manufacture's instruction in the AMV Reverse Transcriptase <sup>35</sup>S-ATP sequencing kit from Stratagene Cloning Systems, La Jolla, CA. The sequence of the resulting V<sub>H</sub> expression vector is shown in Figure 6A and Figure 7.

**SUBSTITUTE SHEET**

Table 9

	(91) N1)	5' GGCCGCAAATTCTATTTCAAGGAGACAGTCAT 3'
	(92) N2)	5' AATGAAATACCTATTTGCCTACGGCAGCCGCTGGATT 3'
	(93) N3)	5' GTTATTACTCGCTGCCCAACCAGCCATGGCCC 3'
5	(94) N4)	5' AGGTGAAACTGCTCGAGAATTCTAGACTAGGTTAATAG 3'
	(95) N5)	5' TCGACTATTAACTAGTCTAGAATTCTCGAG 3'
	(96) N6)	5' CAGTTTCACCTGGGCCATGGCTGGTTGGG 3'
	(97) N7)	5' CAGCGAGTAATAACAATCCAGCGGCTGCCGTAGGCAATAG 3'
	(98) N8)	5' GTATTTTCATTATGACTGTCTCCTTGAAATAGAATTTGC 3'
10	(99) N9-4)	5' AGGTGAAACTGCTCGAGATTTCTAGACTAGTTACCCGTAC 3'
	(100) N11)	5' GACGTTCCGGACTACGGTTCTTAATAGAATTCG 3'
	(101) N12)	5' TCGACGAATTCTATTAAGAACCGTAGTC 3'
	(102) N10-5)	5' CGGAACGTCGTACGGGTAAGTCTAGAAATCTCGAG 3'

#### 11. V<sub>L</sub> Expression Vector Construction

15 To express the plurality of V<sub>L</sub> coding polynucleotides in an E. coli host cell, a vector was constructed that placed the V<sub>L</sub> coding polynucleotide in the proper reading frame, provided a ribosome binding site as described by Shine et al., Nature, 254:34, (1975), provided a leader  
 20 sequence directing the expressed protein to the periplasmic space and also provided a polynucleotide that coded for a spacer protein between the V<sub>L</sub> polynucleotide and the polynucleotide coding for the epitope tag. A  
 25 synthetic DNA sequence containing all of the above polynucleotides and features was constructed by designing single stranded polynucleotide segments of 20-40 bases that would hybridize to each other and form the double  
 stranded synthetic DNA sequence shown in Figure 6B. The individual single-stranded polynucleotides (N1-N8) are  
 30 shown in Table 9.

Polynucleotides N2, N3, N4, N6, N7 and N8 were kinased by adding 1 ul of each polynucleotide and 20 units of T4 polynucleotide kinase to a solution containing 70 mM Tris-HCL at pH 7.6, 10 mM MgCl<sub>2</sub>, 5 mM DTT, 10 mM 2ME, 500  
 35 micrograms per ml of BSA. The solution was maintained at 37°C for 30 minutes and the reaction stopped by maintain-

**SUBSTITUTE SHEET**



ing the solution at 65°C for 10 minutes. The two end polynucleotides 20 ng of polynucleotides N1 and polynucleotides N5 were added to the above kinasing reaction solution together with 1/10 volume of a solution containing 20.0 mM Tris-HCL at pH 7.4, 2.0 mM MgCl<sub>2</sub> and 50.0 mM NaCl. This solution was heated to 70°C for 5 minutes and allowed to cool to room temperature, approximately 25°C, over 1.5 hours in a 500 ml beaker of water. During this time period all the polynucleotides annealed to form the double stranded synthetic DNA insert. The individual polynucleotides were covalently linked to each other to stabilize the synthetic DNA insert with adding 40 ul of the above reaction to a solution containing 50 ul Tris-HCL at pH 7.5, 7 mM MgCl<sub>2</sub>, 1 mM DTT, 1 mM ATP and 10 units to T4 DNA ligase. This solution was maintained at 37°C for 30 minutes and then the T4 DNA ligase was inactivated by maintaining the solution at 65°C for 10 minutes. The end polynucleotides were kinased by mixing 52 ul of the above reaction, 4 ul of a solution recontaining 10 mM ATP and 5 units of T4 polynucleotide kinase. This solution was maintained at 37°C for 30 minutes and then the T4 polynucleotide kinase was inactivated by maintaining the solution at 65°C for 10 minutes. The completed synthetic DNA insert was ligated directly into a lambda Zap II vector that had been previously digested with the restriction enzymes Not I and Xho I. The ligation mixture was packaged according to the manufacture's instructions using Gigapack II Gold packing extract available from Stratagene Cloning Systems, La Jolla, CA. The packaged ligation mixture was plated on XL1-Blue cells (Stratagene Cloning Systems, La Jolla, CA). Individual lambda Zap II plaques were cored and the inserts excised according to the *in vivo* excision protocol provided by the manufacturer, Stratagene Cloning Systems, La Jolla, CA and described in Short et al., Nucleic Acids Res., 16:7583-7600 (1988). This *in vivo* excision protocol moves the cloned insert from the lambda Zap II vector into a phagemid vector to

**SUBSTITUTE SHEET**

allow easy manipulation and sequencing and also produces the phagemid version of the  $V_L$  expression vectors. The accuracy of the above cloning steps was confirmed by sequencing the insert using the Sanger dideoxide method described by Sanger et al., Proc. Natl. Acad. Sci. USA, 74:5463-5467, (1977) and using the manufacturer's instructions in the AMV reverse transcriptase  $^{35}\text{S}$ -dATP sequencing kit from Stratagene Cloning Systems, La Jolla, CA. The sequence of the resulting  $V_L$  expression vector is shown in Figure 6 and Figure 8.

The  $V_L$  expression vector used to construct the  $V_L$  library was the phagemid produced to allow the DNA of the  $V_L$  expression vector to be determined. The phagemid was produced, as detailed above, by the in vivo excision process from the Lambda Zap  $V_L$  expression vector (Figure 8). The phagemid version of this vector was used because the Nco I restriction enzyme site is unique in this version and thus could be used to operatively linked the  $V_L$  DNA homologs into the expression vector.

## 12. $V_{LII}$ -Expression Vector Construction

To express the plurality of  $V_L$ -coding DNA homologs in an *E. coli* host cell, a vector was constructed that placed the  $V_L$ -coding DNA homologs in the proper reading frame, provided a ribosome binding site as described by Shine et al., Nature, 254:34, 1975, provided the Pel B gene leader sequence that has been previously used to successfully secrete Fab fragments in *E. coli* by Lei et al., J. Bac., 169:4379 (1987) and Better et al., Science, 240:1041 (1988), and also provided a polynucleotide containing a restriction endonuclease site for cloning. A synthetic DNA sequence containing all of the above polynucleotides and features was constructed by designing single stranded polynucleotide segments of 20-60 bases that would hybridize to each other and form the double stranded synthetic DNA sequence shown in Figure 10. The sequence of each individual single-stranded polynucleotides (01-08) within

the double stranded synthetic DNA sequence is shown in Table 10.

Polynucleotides 02, 03, 04, 05, 06 and 07 were kinased by adding 1 ul (0.1 ug/ul) of each polynucleotide and 20 units of T4 polynucleotide kinase to a solution containing 70 mM Tris-HCL at pH 7.6, 10 mM magnesium chloride (MgCl), 5 mM dithiothreitol (DTT), 10 mM 2-mercaptoethanol (2ME), 500 micrograms per ml of bovine serum albumin. The solution was maintained at 37°C for 30 minutes and the reaction stopped by maintaining the solution at 65°C for 10 minutes. The 20 ng each of the two end polynucleotides, 01 and 08, were added to the above kinasing reaction solution together with 1/10 volume of a solution containing 20.0 mM Tris-HCl at pH 7.4, 2.0 mM MgCl and 15.0 mM sodium chloride (NaCl). This solution was heated to 70°C for 5 minutes and allowed to cool to room temperature, approximately 25°C, over 1.5 hours in a 500 ml beaker of water. During this time period all 8 polynucleotides annealed to form the double stranded synthetic DNA insert shown in Figure 9. The individual polynucleotides were covalently linked to each other to stabilize the synthetic DNA insert by adding 40 ul of the above reaction to a solution containing 50 ml Tris-HCl at pH 7.5, 7 ml MgCl, 1 mm DTT, 1 mm ATP and 10 units of T4 DNA ligase. This solution was maintained at 37°C for 30 minutes and then the T4 DNA ligase was inactivated by maintaining the solution at 65°C for 10 minutes. The end polynucleotides were kinased by mixing 52 ul of the above reaction, 4 ul of a solution containing 10 mM ATP and 5 units of T4 polynucleotide kinase. This solution was maintained at 37°C for 30 minutes and then the T4 polynucleotide kinase was inactivated by maintaining the solution at 65°C for 10 minutes. The completed synthetic DNA insert was ligated directly into a lambda Zap II vector that had been previously digested with the restriction enzymes Not I and Xho I. The ligation mixture was packaged according to the manufacturer's instructions

**SUBSTITUTE SHEET**

using Gigapack II Gold packing extract available from Stratagene Cloning Systems, La Jolla, CA. The packaged ligation mixture was plated on XL1 blue cells (Stratagene Cloning Systems, San Diego, CA). Individual lambda Zap II

5 plaques were cored and the inserts excised according to the in vivo excision protocol provided by the manufacturer, Stratagene Cloning Systems, La Jolla, CA. This in vivo excision protocol moves the cloned insert from the lambda Zap II vector into a plasmid vector to allow easy

10 manipulation and sequencing. The accuracy of the above cloning steps was confirmed by sequencing the insert using the manufacturer's instructions in the AMV Reverse Transcriptase <sup>35</sup>S-dATP sequencing kit from Stratagene Cloning Systems, La Jolla, CA. The sequence of the

15 resulting V<sub>L</sub>II-expression vector is shown in Figure 9 and Figure 11.

Table 10

	(102) 01) 5'	TGAATTCTAAACTAGTCGCCAAGGAGACAGTCAT 3'
	(103) 02) 5'	AATGAAATACCTATTGCCTACGGCAGCCGCTGGATT 3'
20	(104) 03) 5'	GTTATTACTCGCTGCCCAACCAGCCATGGCC 3'
	(105) 04) 5'	GAGCTCGTCAGTTCTAGAGTTAAGCGGCCG 3'
	(106) 05) 5'	
		GTATTTTCATTATGACTGTCTCCTTGGCGACTAGTTTAGAATTCAAGCT
		3'
25	(107) 06) 5'	CAGCGAGTAATAACAATCCAGCGGCTGCCGTAGGCAATAG 3'
	(108) 07) 5'	TGACGAGCTCGGCCATGGCTGGTTGGG 3'
	(109) 08) 5'	TCGACGGCCGCTTAACTCTAGAAC 3'

### 13. V<sub>H</sub> + V<sub>L</sub> Library Construction

To prepare an expression library enriched in V<sub>H</sub> sequences, DNA homologs enriched in V<sub>H</sub> sequences were prepared according to Example 7 using the same set of 5' primers but with primer 62A (Table 7) as the 3' primer. These homologs were then digested with the restriction enzymes Xho I and Spe I and purified on a 1% agarose gel

35 using the standard electroelution technique described in

**SUBSTITUTE SHEET**

Molecular Cloning A Laboratory Manual, Maniatis et al., eds., Cold Spring Harbor, New York, (1982). These prepared  $V_H$  DNA homologs were then directly inserted into the  $V_H$  expression vector that had been previously digested  
5 with Xho I and Spe I.

The ligation mixture containing the  $V_H$  DNA homologs were packaged according to the manufacturers specifications using Gigapack Gold II Packing Extract (Stratagene Cloning Systems, La Jolla, CA). The expression libraries  
10 were then ready to be plated on XL-1 Blue cells.

To prepare a library enriched in  $V_L$  sequences, PCR amplified products enriched in  $V_L$  sequences were prepared according to Example 7. The  $V_L$  DNA homologs were digested with restriction enzymes Nco I and Spe I. The digested  $V_L$   
15 DNA homologs were purified on a 1% agarose gel using standard electrophoresis techniques described in Molecular Cloning A Laboratory Manual, Maniatis et al., eds., Cold Spring Harbor, NY (1982). The prepared  $V_L$  DNA homologs were directly inserted into the  $V_L$  expression vector that  
20 had been previously digested with the restriction enzymes Nco I and Spe I. The ligation mixture containing the  $V_L$  DNA homologs were transformed into XL-1 blue competent cells using the manufacturer's instructions (Stratagene Cloning Systems, La Jolla, CA).

25 14. Inserting  $V_L$  Coding DNA Homologs Into  $V_L$  Expression Vector

In preparation for cloning a library enriched in  $V_L$  sequences, PCR amplified products (2.5 ug/30 ul of 150 mM NaCl, 8 mM Tris-HCl (pH 7.5), 6 mM MgCl<sub>2</sub>, 1 mM DTT, 200  
30 ug/ml BSA at 37°C were digested with restriction enzymes Sac I (125 units) and Xba I (125 units) and purified on a 1% agarose gel. In cloning experiments which required a mixture of the products of the amplification reactions, equal volumes (50 ul, 1-10 ug concentration) of each  
35 reaction mixture were combined after amplification but before restriction digestion. After gel electrophoresis

of the digested PCR amplified spleen mRNA, the region of the gel containing DNA fragments of approximately 350 bps was excised, electroeluted into a dialysis membrane, ethanol precipitated and resuspended in a TE solution containing 10 mM Tris-HCl pH 7.5 and 1 mM EDTA to a final concentration of 50 ng/ul.

The  $V_L$ II-expression DNA vector was prepared for cloning by admixing 100 ug of this DNA to a solution containing 250 units each of the restriction endonucleases Sac 1 and Xba 1 (both from Boehringer Mannheim, Indianapolis, IN) and a buffer recommended by the manufacturer. This solution was maintained at 37°C for 1.5 hours. The solution was heated at 65°C for 15 minutes to inactivate the restriction endonucleases. The solution was chilled to 30°C and 25 units of heat-killable (HK) phosphatase (Epicenter, Madison, WI) and  $CaCl_2$  were admixed to it according to the manufacturer's specifications. This solution was maintained at 30°C for 1 hour. The DNA was purified by extracting the solution with a mixture of phenol and chloroform followed by ethanol precipitation. The  $V_L$ II expression vector was now ready for ligation to the  $V_L$  DNA homologs prepared in the above examples.

DNA homolog enriched in  $V_L$  sequences were prepared according to Example 6 but using a 5' light chain primer and 3' light chain primer shown in Table 9. Individual amplification reactions were carried out using each 5' light chain primer in combination with the 3' light chain primer. These separate  $V_L$  homolog-containing reaction mixtures were mixed and digested with the restriction endonucleases Sac 1 and Xba 1 according to Example 7. The  $V_L$  homologs were purified on a 1% agarose gel using the standard electroelution technique described in Molecular Cloning A Laboratory Manual, Maniatis et al., eds., Cold Spring Harbor, New York, (1982). These prepared  $V_L$  DNA homologs were then directly inserted into the Sac 1 - Xba cleaved  $V_L$ II-expression vector that was prepared above by ligating 3 moles of  $V_L$  DNA homolog inserts with each mole

of the  $V_{II}$ -expression vector overnight at 5°C.  $3.0 \times 10^5$  plaque forming units were obtained after packaging the DNA with Gigapack II Bold (Stratagene Cloning Systems, La Jolla, CA) and 50% were recombinants.

5 15. Randomly Combining  $V_H$  and  $V_L$  DNA Homologs on the Same Expression Vector

The  $V_{II}$ -expression library prepared in Example 13 was amplified and 500 ug of  $V_{II}$ -expression library phage DNA prepared from the amplified phage stock using the procedures described in Molecular Cloning: A Laboratory Manual, Maniatis et al., eds., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY (1982), 50 ug of this  $V_{II}$ -expression library phage DNA was maintained in a solution containing 100 units of  $LuI$  restriction endo-  
10 nuclease (Boehringer Mannheim, Indianapolis, IN) in 200 ul of a buffer supplied by the endonuclease manufacturer for 1.5 hours at 37°C. The solution was then extracted with a mixture of phenol and chloroform. The DNA was then ethanol precipitated and resuspended in 100 ul of water.  
15 This solution was admixed with 100 units of the restriction endonuclease  $EcoR$  I (Boehringer Mannheim, Indianapolis, IN) in a final volume of 200 ul of buffer containing the components specified by the manufacturer. This solution was maintained at 37°C for 1.5 hours and the  
20 solution was then extracted with a mixture of phenol and chloroform. The DNA was ethanol precipitated and the DNA resuspended in TE.

The  $V_H$  expression library prepared in Example 13 was amplified and 500 ug of  $V_H$  expression library phage DNA  
30 prepared using the methods detailed above. 50 ug of the  $V_H$  expression library phage DNA was maintained in a solution containing 100 units of  $Hind$  III restriction endonuclease (Boehringer Mannheim, Indianapolis, IN) in 200 ul of a buffer supplied by the endonuclease manufacturer for  
35 1.5 hours at 37°C. The solution was then extracted with a mixture of phenol and chloroform saturated with 0.1

Tris-HCL at pH 7.5. The DNA was then ethanol precipitated and resuspended in 100 ul of water. This solution was admixed with 100 units of the restriction endonuclease EcoR I (Boehringer Mannheim, Indianapolis, IN) in a final  
5 volume of 200 ul of buffer containing the components specified by the manufacturer. This solution was maintained at 37°C for 1.5 hours and the solution was then extracted with a mixture of phenol and chloroform. The DNA was ethanol precipitated and the DNA resuspended in  
10 TE.

The restriction digested  $V_H$  and  $V_L$ II-expression Libraries were ligated together. The ligation reaction consisted of 1 ug of  $V_H$  and 1 ug of  $V_L$ II phage library DNA in a 10 ul reaction using the reagents supplied in a ligation  
15 kit purchased from Stratagene Cloning Systems (La Jolla, CA). After ligation for 16 hr at 4°C, 1 ul of the ligated phage DNA was packaged with Gigapack Gold II packaging extract and plated on XL 1-blue cells prepared according the manufacturer's instructions. A portion of  
20 the  $3 \times 10^6$  clones obtained were used to determine the effectiveness of the combination. The resulting  $V_H$  and  $V_L$  expression vector is shown in Figure 11.

Clones containing both  $V_H$  and  $V_L$  were excised from the phage to pBluescript using the in vitro excision protocol  
25 described by Short et al., Nucleic Acid Research, 16L7583-7600 (1988). Clones chosen for excision expressed the decapeptide tag and did not cleave X-gal in the presence of 2mM IPTG, thus remaining white. Clones with these characteristics represented 30% of the library. 50% of  
30 the clones chosen for excision contained a  $V_H$  and  $V_L$  as determined by restriction analysis. Since approximately 30% of the clones in the  $V_H$  library expressed the decapeptide tag and 50% of the clones in the  $V_L$ II library contained a  $V_L$  sequence it was anticipated that no more  
35 than 15% of the clones in the combined library would contain both  $V_H$  and  $V_L$  clones. The actual number obtained

**SUBSTITUTE SHEET**



was 15% of the library indicating that the process of combination was very efficient.

16. Segregating DNA Homologs For a  $V_H$  Antigen Binding Protein

5 To segregate the individual clones containing DNA homologs that code for a  $V_H$  antigen binding protein, the titre of the  $V_H$  expression library prepared according to Example 12 was determined. This library titration was performed using methods well known to one skilled in the  
10 art. Briefly, serial dilutions of the library were made into a buffer containing 100 mM NaCl, 50 mM Tris-HCL at pH 7.5 and 10 mM  $MgCl_4$ , 5 g/L yeast extract, 10 g/L NZ amine (casein hydrolysate) and 0.7% melted, 50°C agarose. The phage, the bacteria and the top agar were mixed and  
15 then evenly distributed across the surface of a prewarmed bacterial agar plate (5 g/L NaCl, 2 g/L  $MgCl_4$ , 5 g/L yeast extract, 10 g/L NZ amine (casein hydrolysate) and 15 g/L Difco agar. The plates were maintained at 37°C for 12 to 24 hours during which time period the lambda plaques  
20 developed on the bacterial lawn. The lambda plaques were counted to determine the total number of plaque forming units per ml in the original library.

The titred expression library was then plated out so that replica filters could be made from the library. The  
25 replica filters will be used to later segregate out the individual clones in the library that are expressing the antigens binding proteins of interest. Briefly, a volume of the titred library that would yield 20,000 plaques per 150 millimeter plate was added to 600 ul of exponentially  
30 growing E. coli cells and maintained at 37°C for 15 minutes to allow the phage to absorb to the bacterial cells. The 7.5 ml of top agar was admixed to the solution containing the bacterial cells and the absorbed phage and the entire mixture distributed evenly across the surface  
35 of a prewarmed bacterial agar plate. This process was repeated for a sufficient number of plates to plate out a

total number of plaques at least equal to the library size. These plates were then maintained at 37°C for 5 hours. The plates were then overlaid with nitrocellulose filters that had been pretreated with a solution containing 10 mM isopropyl-beta-D-thiogalactopyranoside (IPTG) and maintained at 37°C for 4 hours. The orientation of the nitrocellulose filters in relation to the plate were marked by punching a hole with a needle dipped in waterproof ink through the filter and into the bacterial plates at several locations. The nitrocellulose filters were removed with forceps and washed once in a TBST solution containing 20 mM Tris-HCl at pH 7.5, 150 mM NaCl and 0.05% monolaurate (tween-20). A second nitrocellulose filter that had also been soaked in a solution containing 10 mM IPTG was reapplied to the bacterial plates to produce duplicate filters. The filters were further washed in a fresh solution of TBST for 15 minutes. Filters were then placed in a blocking solution consisting 20 mM Tris-HCl at pH 7.5, 150 mM NaCl and 1% BSA and agitated for 1 hour at room temperature. The nitrocellulose filters were transferred to a fresh blocking solution containing a 1 to 500 dilution of the primary antibody and gently agitated for at least 1 hour at room temperature. After the filters were agitated in the solution containing the primary antibody the filters were washed 3 to 5 times in TBST for 5 minutes each time to remove any of the residual unbound primary antibody. The filters were transferred into a solution containing fresh blocking solution and a 1 to 500 to a 1 to 1,000 dilution of alkaline phosphatase conjugated secondary antibody. The filters were gently agitated in the solution for at least 1 hour at room temperature. The filters were washed 3 to 5 times in a solution of TBST for at least 5 minutes each time to remove any residual unbound secondary antibody. The filters were washed once in a solution containing 20 mM Tris-HCl at pH 7.5 and 150 mM NaCl. The filters were removed from this solution and excess moisture blotted

**SUBSTITUTE SHEET**

from them with filter paper. The color was developed by placing the filter in a solution containing 100 mM Tris-HCl at pH 9.5, 100 mM NaCl, 5 mM MgCl<sub>2</sub>, 0.3 mg/ml of nitro Blue Tetrazolium (NBT) and 0.15 mg/ml of 5-bromo-4-chloro-3-indolyl-phosphate (BCIP) for at least 30 minutes at room temperature. The residual color development solution was rinsed from the filter with a solution containing 20 mM Tris-HCl at pH 7.5 and 150 mM NaCl. The filter was then placed in a stop solution consisting of 20 mM Tris-HCl at pH 2.9 and 1 mM EDTA. The development of an intense purple color indicates a positive result. The filters are used to locate the phage plaque that produced the desired protein. That phage plaque is segregated and then grown up for further analysis.

Several different combinations of primary antibodies and second antibodies were used. The first combination used a primary antibody immunospecific for a decapeptide that will be expressed only if the V<sub>H</sub> antigen binding protein is expressed in the proper reading frame to allow read through translation to include the decapeptide epitope covalently attached to the V<sub>H</sub> antigen binding protein. This decapeptide epitope and an antibody immunospecific for this decapeptide epitope was described by Green et al., Cell 28:477 (1982) and Niman et al., Proc. Nat. Acad. Sci. U.S.A. 80:4949 (1983). The sequence of the decapeptide recognized is shown in Figure 11. A functional equivalent of the monoclonal antibody that is immunospecific for the decapeptide can be prepared according to the methods of Green et al. and Niman et al. The secondary antibody used with this primary antibody was a goat antimouse IgG (Fisher Scientific). This antibody is immunospecific for the strand region of mouse IgG and did not recognize any portion of the variable region of heavy chain. This particular combination of primary and secondary antibodies when used according to the above protocol determined that between 25% and 30% of the clones were

expressing the decapeptide and therefore these clones were assumed to also be expressing a  $V_H$  antigen binding protein.

In another combination the anti-decapeptide mouse monoclonal was used as the primary antibody and an affinity purified goat anti-mouse Ig, commercially available as part of the picoBlue immunoscreening kit from Stratagene Cloning System, La Jolla, CA, was used as the secondary antibody. This combination resulted in a large number of false positive clones because the secondary antibody also immunoreacted with the  $V_H$  of the heavy chain. Therefore this antibody reacted with all clones expressing any  $V_H$  protein and this combination of primary and secondary antibodies did not specifically detect clones with the  $V_H$  polynucleotide in the proper reading frame and thus allowing expressing of the decapeptide.

Several combinations of primary and secondary antibodies are used where the primary antibody is conjugated to fluorescein isothiocyanate (FITC) and thus the immunospecificity of the antibody was not important because the antibody is conjugated to the preselected antigen (FITC) and it is that antigen that should be bound by the  $V_H$  antigen binding proteins produced by the clones in the expression library. After this primary antibody has bound by virtue that is FITC conjugated mouse monoclonal antibody p2 5764 (ATCC #HB-9505). The secondary antibody used with this primary antibody is a goat anti-mouse Ig<sup>6</sup> (Fisher Scientific, Pittsburgh, PA) conjugated to alkaline phosphatase using the method described in Antibodies: A Laboratory Manual, Harlow and Lowe, eds., Cold Spring Harbor, New York, (1988). If a particular clone in the  $V_H$  expression library, expresses a  $V_H$  binding protein that binds the FITC covalently coupled to the primary antibody, the secondary antibody binds specifically and when developed the alkaline phosphate causes a distinct purple color to form.

The second combination of antibodies of the type uses a primary antibody that is FITC conjugated rabbit anti-

human IgG (Fisher Scientific, Pittsburgh, PA). The secondary antibody used with this primary antibody is a goat anti-rabbit IgG conjugated to alkaline phosphatase using the methods described in Antibodies A Laboratory Manual,  
5 Harlow and Lane, eds., Cold Spring Harbor, New York, (1988). If a particular clone in the  $V_H$  expression library expresses a  $V_H$  binding protein that binds the FITC conjugated to the primary antibody, the secondary antibody binds specifically and when developed the alkaline  
10 phosphatase causes a distinct purple color to form.

Another primary antibody was the mouse monoclonal antibody p2 5764 (ATCC # HB-9505) conjugated to both FITC and  $^{125}\text{I}$ . The antibody would be bound by any  $V_H$  antigen binding proteins expressed. Then because the antibody is  
15 also labeled with  $^{125}\text{I}$ , an autoradiogram of the filter is made instead of using a secondary antibody that is conjugated to alkaline phosphatase. This direct production of an autoradiogram allows segregation of the clones in the library expressing a  $V_H$  antigen binding protein of  
20 interest.

17. Segregating DNA Homologs For a  $V_H$  and  $V_L$  that Form an Antigen Binding  $F_v$

To segregate the individual clones containing DNA homologs that code for a  $V_H$  and  $V_L$  that form an antigen  
25 binding  $F_v$ , an  $V_H$  and  $V_L$  expression library was titred according to Example 15. The titred expression library was then screened for the presence of the decapeptide tag expressed with the  $V_H$  using the methods described in Example 16. DNA was then prepared from the clones to  
30 express the decapeptide tag. This DNA was digested with the restriction endonuclease Pvu II to determine whether these clones also contained a  $V_L$  DNA homolog. The slower migration of a PvuII restriction endonuclease fragment indicated that the particular clone contained both a  $V_H$  and  
35 a  $V_L$  DNA homolog.

The clones containing both a  $V_H$  and a  $V_L$  DNA homolog were analyzed to determine whether these clones produced an assembled  $F_v$  protein molecule from the  $V_H$  and  $V_L$  DNA homologs.

- 5        The  $F_v$  protein fragment produced in clones containing both  $V_H$  and  $V_L$  was visualized by immune precipitation of radiolabeled protein expressed in the clones. A 50 ml culture of LB broth (5 g/L yeast extract, 10 g/L and tryptone 10 g/L NaCl at pH 7.0) containing 100 ug/ul of  
10    ampicillin was inoculated with E. Coli harboring a plasmid contain a  $V_H$  and a  $V_L$ . The culture was maintained at 37°C with shaking until the optical density measured at 550 nm was 0.5. The culture then was centrifuged at 3,000 g for 10 minutes and resuspended in 50 ml of M9 media (6 g/L  
15     $Na_2HPO_4$ , 3 g/L  $KH_2PO_4$ , 0.5 g/L NaCl, 1 g/L  $NH_4Cl$ , 2g/L glucose, 2 mM  $MgSO_4$  and 0.1 mM  $MgSO_4$   $CaCl_2$  supplemented with amino acids without methionine or cysteine. This solution was maintained at 37°C for 5 minutes and then 0.5 mCi of  $^{35}S$  as  $HSO_4$  (New England Nuclear, Boston, MA) was added and  
20    the solution was further maintained at #&C for an additional 2 hours. The solution was then centrifuged at 300 x g and the supernatant discarded. The resulting bacterial cell pellet was frozen and thawed and then resuspended for 10 minutes and the resulting pellet discarded. The  
25    supernatant was admixed with 10 ul of anti-decapeptide monoclonal antibody and maintained for 30-90 minutes on ice. 40 ul of protein G coupled to sepharose beads (Pharmacia, Piscataway, NJ) was admixed to the solution and the added solution maintained for 30 minutes on ice to  
30    allow an immune precipitate to form. The solution was centrifuged at 10,000 x g for 10 minutes and the resulting pellet was resuspended in 1 ml of a solution containing 100 mM Tris-HCl at Ph 7.5 and centrifuged at 10,000 x g for 10 minutes. This procedure was repeated twice. The  
35    resulting immune precipitate pellet was loaded onto a PhastGel Homogenous 20 gel (Pharmacia, Piscataway, NJ)

according to the manufacturer's directions. The gel was dried and used to expose X-ray film.

The resulting autoradiogram is shown in Figure 12. The presence of  $V_L$  that was immunoprecipitated because it was attached to the  $V_H$ -decapeptide tag recognized by the precipitating antibody.

#### 18. Generation of a Combinatorial Library of the Immunoglobulin Repertoire in Phage

Vectors suitable for expression of  $V_H$ ,  $V_L$ , Fv and Fab sequences are diagrammed in Figures 7 and 9. As previously discussed, the vectors were constructed by modification of Lambda Zap by inserting synthetic oligonucleotides into the multiple cloning site. The vectors were designed to be antisymmetric with respect to the Not I and EcoR I restriction sites which flank the cloning and expression sequences. As described below, this antisymmetry in the placement of restriction sites in a linear vector like bacteriophage allows a library expressing light chains to be combined with one expressing heavy chains to construct combinatorial Fab expression libraries. Lambda Zap II  $V_L$ II (Figure 9) is designed to serve as a cloning vector for light chain fragments and Lambda Zap II  $V_H$  (Figure 7) is designed to serve as a cloning vector for heavy chain sequences in the initial step of library construction. These vectors are engineered to efficiently clone the products of PCR amplification with specific restriction sites incorporated at each end.

##### A. PCR Amplification of Antibody Fragments

The PCR amplification of mRNA isolated from spleen cells with oligonucleotides which incorporate restriction sites into the ends of the amplified product can be used to clone and express heavy chain sequences including Fd and kappa chain sequences. The oligonucleotides primers used for these amplifications are presented in Tables 1

and 2. The primers are analogous to those which have been successfully used in Example 6 for amplifications of  $V_H$  sequences. The set of 5' primers for heavy chain amplification were identical to those previously used to amplify  $V_H$  and those for light chain amplification were chosen on similar principles, Sastry et al., Proc. Natl. Acad. Sci. USA, 86: 5728 (1989) and Orland et al., Proc Natl. Acad. Sci. USA, 86:3833 (1989). The unique 3' primers of heavy (IgG1) and light (k) chain sequences were chosen to include the cysteines involved in heavy-light chain disulfide bond formation. At this stage no primer was constructed to amplify lambda light chains since they constitute only a small fraction of murine antibodies. In addition, Fv fragments have been constructed using a 3' primer which is complementary to the to the mRNA in the J (joining) region (amino acid 128) and a set of unique 5' primers which are complementary to the first strand cDNA in the conserved N-terminal region of the processed protein. Restriction endonuclease recognition sequences are incorporated into the primers to allow for the cloning of the amplified fragment into a lambda phage vector in a predetermined reading frame for expression.

#### B. Library Construction

The construction of a combinatorial library was accomplished in two steps. In the first step, separate heavy and light chain libraries were constructed in Lambda Zap II  $V_H$  and Lambda Zap II  $V_L$  II respectively. In the second step, these two libraries were combined at the antisymmetric EcoRI sites present in each vector. This resulted in a library of clones each of which potentially co-expresses a heavy and a light chain. The actual combinations are random and do not necessarily reflect the combinations present in the B-cell population in the parent animal. Lambda Zap II  $V_H$  expression vector has been used to create a library of heavy chain sequences for DNA obtained by PCR amplifications of mRNA isolated from the

**SUBSTITUTE SHEET**



spleen of a 129 G<sub>ix</sub> + mouse previously immunized with p-nitrophenyl phosphoramidate (NPN) antigen 1 according to formula I (Figure 13) conjugated to keyhole limpet hemocyanin (KLH).

5       The NPN-KLH conjugate was prepared by admixture of 250 ul of a solution containing 2.5 mg of NPN according to formula 1 (Figure 12) in dimethylformamide with 750 ul of a solution containing 2 mg of KLH in 0.01 M sodium phosphate buffer (pH 7.2). The two solutions were admixed by  
10 slow addition of the NPN solution to the KLH solution while the KLH solution was being agitated by a rotating stirring bar. Thereafter the admixture was maintained at 4°C for 1 hour with the same agitation to allow conjugation to proceed. The conjugated NPN-KHL was isolated from  
15 the nonconjugated NPN and KLH by gel filtration through Sephadex G-25. The isolated NPN-KLH conjugate was used in mouse immunizations as described in Example 3.

The spleen mRNA resulting from the above immunizations was isolated and used to create a primary library of  
20 V<sub>H</sub> gene sequences using the Lambda Zap II V<sub>H</sub> expression vector. The primary library contains  $1.3 \times 10^6$  pfu and has been screened for the expression of the decapeptide tag to determine the percentage of clones expressing Fd sequences. The sequence for this peptide is only in frame  
25 for expression following the cloning of an Fd (or V<sub>H</sub>) fragment into the vector. At least 80% of the clones in the library express Fd fragments based on immuno-detection of the decapeptide tag.

The light chain library was constructed in the same  
30 way as the heavy chain and shown to contain  $2.5 \times 10^6$  members. Plaque screening, using the anti-kappa chain antibody, indicated that 60% of the library contained expressed light chain inserts. This relatively small percentage of inserts probably resulted from incomplete  
35 dephosphorylation of vector after cleavage with Sac I and Xba I.

Once obtained, the two libraries were used to construct a combinatorial library by crossing them at the EcoRI site. To accomplish the cross, DNA was first purified from each library. The light chain library was  
5 cleaved with MluI restriction endonuclease, the resulting 5' ends dephosphorylated and the product digested with EcoRI. This process cleaved the left arm of the vector into several pieces but the light arm containing the light chain sequences, remained intact. In a parallel fashion,  
10 the DNA of heavy chain library was cleaved with HindIII, dephosphorylated and cleaved with EcoR I, destroying the right arm but leaving the left arm containing the heavy chain sequences intact. The DNA's so prepared were then combined and ligated. After ligation only clones which  
15 resulted from combination of a right arm of light chain-containing clones reconstituted a viable phage. After ligation and packaging,  $2.5 \times 10^7$  clones were obtained. This is the combinatorial Fab expression library that was screened to identify clones having affinity for NPN. To  
20 determine the frequency the phage clones which co-express the light and heavy chain fragments, duplicate lifts of the light chain, heavy chain and combinatorial libraries were screened as above for light and heavy chain expression. In this study of approximately 500 recombinant phage  
25 approximately 60% co-expressed light and heavy chain proteins.

### C. Antigen Binding

All three libraries, the light chain, the heavy chain and Fab were screened to determine if they contained  
30 recombinant phage that expressed antibody fragments binding NPN. In a typical procedure 30,000 phage were plated and duplicate lifts with nitrocellulose screened for binding to NPN coupled to  $^{125}\text{I}$  labeled BSA (Figure 15). Duplicate screens of 80,000 recombinant phage from the  
35 light chain library and a similar number from the heavy chain library did not identify any clones which bound the

antigen. In contrast, the screen of a similar number of clones from the Fab expression library identified many phage plaques that bound NPN (Figure 15). This observation indicates that under conditions where many heavy chains in combination with light chains bind to antigen the same heavy or light chains alone do not. Therefore, in the case of NPN, it is believed that there are many heavy and light chains that only bind antigen when they are combined with specific light and heavy chains respectively.

To assess the ability to screen large numbers of clones and obtain a more quantitative estimate of the frequency of antigen binding clones in the combinatorial library, one million phage plaques were screened and approximately 100 clones which bound to antigen were identified. For six clones which were believed to bind NPN, a region of the plate containing the positive and approximately 20 surrounding bacteriophage plaques was "cored", replated, and screened with duplicate lifts (Figures 15). As expected, approximately one in twenty of the phage specifically bind to antigen. "Cores" of regions of the plated phage believed to be negative did not give positives on replating.

To determine the specificity of the antigen-antibody interaction, antigen binding was competed with free unlabeled antigen as shown in Figure 16. Competition studies showed that individual clones could be distinguished on the basis of antigen affinity. The concentration of free hapten required for complete inhibition of binding varied between  $10-100 \times 10^{-9}$  M suggesting that the expressed Fab fragments had binding constants in the nanomolar range.

#### D. Composition of the Clones and Their Expressed Products

In preparation for characterization of the protein products able to bind NPN as described in Example 19C, a

plasmid containing the heavy and light chain genes was excised from the appropriate "cored" bacteriophage plaque using M13mp8 helper phage. Mapping of the excised plasmid demonstrated a restriction pattern consistent with incorporation of heavy and light chain sequences. The protein products of one of the clones was analyzed by ELISA and Western blotting to establish the composition of the NPN binding protein. A bacterial supernate following IPTG induction was concentrated and subjected to gel filtration. Fractions in the molecular weight range 40-60 kD were pooled, concentrated and subjected to a further gel filtration separation. As illustrated in Figure 17, ELISA analysis of the eluting fractions demonstrated that NPN binding was associated with a protein of molecular weight about 50 kD which immunological detection showed contained both heavy and light chains. A Western blot (not shown) of a concentrated bacterial supernate preparation under non-reducing conditions was developed with anti-decapeptide antibody. This revealed a protein band of molecular weight of 50 kD. Taken together these results are consistent with NPN binding being a function of Fab fragments in which heavy and light chains are covalently linked.

#### 20. Flp recombinase-catalyzed Recombination

Experiments directed to the in vivo recombination of two lambda vectors using flp recombinase-catalyzed recombination are described. The flp recombination site was introduced into the phage vectors using 39mer synthetic oligonucleotides. The sequence of the flp site utilized for recombination was derived from several references (e.g. Senecoff et al., Proc. Nat. Acad. Sci. USA 82:7220-7224 (1985)). The XbaI site within the 8bp core was eliminated as this site was to be used in the cloning strategy. This was accomplished by making a point mutation which has little or no effect on its ability to allow recombination (McLeod et al., Mol. Cell. Biol. 6:3357-

3367 (1985)). However, this point mutation is not required for the system to function. The oligonucleotides were further designed to be inserted in the EcoRI sites of the Lambda Zap II V<sub>H</sub> and Lambda Zap II V<sub>L</sub> vectors so that only one flanking EcoRI sites would be regenerated (see Figure 18). The flanking sequences are not essential to the system.

The following sequences were inserted into Lambda Zap II V<sub>H</sub>:

(110) O l i g o 7 9 E c o R I  
AATTCTGAAGTTCCTATTCTCTAAAAAGTATAGGAACTTC 3'  
(111) Oligo 80 GCTTCAAGGATAAGAGATTTTTCATATCCTTGAAGTTAA  
5'

The following sequences were inserted into Lambda Zap II V<sub>L</sub>:

(112) O l i g o 8 1  
AATTGAAGTTCCTATTCTCTAAAAAGTATAGGAACTTCG EcoRI 3'  
(113) Oligo 82 CTTCAAGGATAAGAGATTTTTCATATCCTTGAAGCTTAA  
5'

Vectors were constructed as follows. The first two oligonucleotides were mixed (0.5 µg oligo 79, 0.5 µg oligo 80, 1 µl 200 mM Tris, pH 7.4, 20 mM MgCl, 500 mM NaCl, and H<sub>2</sub>O to 10 µl), heated to 85°C 5 min., and allowed to cool to room temperature over 1 hour in a water bath. The procedure was repeated using oligos 81 and 82.

Ligation into vector arms was accomplished by digesting Lambda Zap V<sub>H</sub> and Lambda Zap II V<sub>L</sub> with 3U/µg EcoRI according to standard digestion procedure. After phenol/chloroform extraction, DNA was precipitated with EtOH. The vector was not phosphatase treated so that the oligonucleotides could be inserted without kinase treatment, thus preventing multiple tandem oligonucleotide inserts. Ligations were performed in 5 µl volumes using 1 µg of lambda DNA and 1 ng of annealed oligonucleotides according to standard ligation protocol (see Maniatis *et al.*, *supra*). Ligation mixes were packaged using Gigapack

Gold<sup>TM</sup> (Stratagene Cloning Systems, San Diego, CA) according to the protocol recommended in the manual.

Following packaging, the vectors were screened. Packaged DNA was plated according to the Gigapack Gold<sup>TM</sup> manual procedure on NZY agar with approximately 400 pfu per 100 mm Petri dish. Duplicate plaque lifts were done according to the protocol in the Predigested ZapII Cloning Manual (Stratagene Cloning Systems, San Diego, CA) on nitrocellulose filters. Denaturation and fixation of DNA onto the membranes is also described in the manual. Prehybridization was performed according to pBluescript II Exo/Mung DNA Sequencing System<sup>TM</sup> instruction manual (Stratagene Cloning Systems, San Diego, CA) for oligo-nucleotide probes (pg 6). Hybridization was performed overnight using <sup>32</sup>P kinased oligo 79 (0.5 x 10<sup>6</sup> cpm/ml) according to the pBluescript manual (Stratagene Cloning Systems, San Diego, CA). Oligo 79 was kinased using standard <sup>32</sup>P gamma ATP labelling techniques (see Maniatis et al., supra). Filters were washed in 6X SSC, 0.1% SDS, 3 times at room temperature, once at 55°C and finally at 59°C. Each was washed for approximately 10 minutes. Positive plaques were identified using X-ray autoradiography. Twelve duplicate plaques were cored in 500 µl SM, 20 µl chloroform. These plaques were sufficiently well isolated that secondary screening was not required. The cored plaques were excised according to the Predigested Zap II Cloning Manual (Stratagene Cloning Systems, San Diego, CA) and DNA from single ampicillin resistant colonies was sequenced using miniprep DNA and the T7 and T3 primers according to the DSK 35S Sequencing kit (Stratagene Cloning Systems, San Diego, CA). Clones with flp sites in the correct orientation and opposite orientation were identified, amplified and titred. One of each type of clone (FLPHC+, FLPHC-, FLPLC+, FLPLC-) was used to test in vivo flp- mediated recombination.

In vivo flp-mediated recombination was accomplished as follows. Flp recombinase was expressed off the tac

**SUBSTITUTE SHEET**

promoter on a plasmid, pCS3, in E. coli MM294 strain (Lebreton et al. 1988). This is a low copy number plasmid with the pACYC origin of replication and contains a chloramphenicol resistance gene.

5         $5 \times 10^8$  cells were coinfectd with FLPHC and FLPLC vectors at an moi of 5 and 10 pfu each per cell. Combinations of FLPHC+ and FLPLV+, or FLPHC- and FLPLC-, or FLPHC+ and FLPLC- were tested.

Overnight cultures of MM294(pCS3) were grown in LB,  
10 spun down and resuspended in 10mM MgSO<sub>4</sub> at a density of OD<sub>600</sub> = 1.0. The appropriate amounts of phage were added to 0.5 ml of cells and allowed to adhere at 37°C for 15 minutes. 50 ml of NZY was added to each flask and incubated for 2 hours with shaking. 250 µl of chloroform was  
15 added to 25 ml of lysate and incubated for 15 minutes at room temperature. The supernatants were titred and screened for phage containing both Lambda Zap II V<sub>H</sub> left arms and Lambda Zap II V<sub>L</sub> right arms. Probes to identify Zap II V<sub>H</sub> left arms and Lambda Zap II V<sub>L</sub> right arms were  
20 designed by identifying unique sequences from the known sequence of the vectors.

The Lambda Zap II V<sub>H</sub> left arm probe had the following sequence:

(114) CTAGTTACCCGTACGACCCCCCGTTCCGGACTACGCTTCTTAATAG 3'

25 This sequence hybridizes to the decapeptide sequence of the Lambda Zap II V<sub>H</sub>. The Lambda Zap II V<sub>L</sub> right arm probe had the following sequence:

(115) 5' GAGCTCGTCAGTTCTAGAGTTAAGCGGCCG 3'

This sequence hybridizes to the sequence from the SacI  
30 site to the former NotI site of the Lambda Zap II V<sub>L</sub> vector.

The screening procedure used was the same as that used to identify the flp vectors, as described above, with the exception of washing conditions. Filters were washed  
35 with 6XSSC, 1%SDS 3 times at room temperature and twice at 60°C. Plaques which hybridized to both probes were identified by X-ray autoradiography, cored, excised and

**SUBSTITUTE SHEET**

digested to determine if recombination had occurred. Control plaques identified as hybridizing to only one probe and to neither probe were also cored. Diagnostic restriction digests were PvuII, PvuI, XhoI, XhoI/PvuI, 5 SacI, Sac/PvuI, NotI, XbaI, ScaI, SpeI, SpeI/PvuI. Restriction digest results verified that recombination at the flp site occurred in vivo in cells expressing the flp recombinase gene and not in control SURE™ E. coli cells (Stratagene Cloning Systems, San Diego, CA) which do not 10 normally express flp recombinase.

Efficiency of recombination according to the number of plaques identified as hybridizing to both probes was initially between about 5-10%. Changes to the protocol can be made, however, which will improve the efficiency of 15 recovery of recombined vectors. For example, by adding selectable marker sequences to the left and right arms of the vectors, up to 100% of target recombinants can be identified (Figure 20). Adding selection systems to ensure that all recombinants contain inserts will also 20 increase the efficiency of identifying the desired clones.

In Example 19 a relatively restricted library was prepared because only a limited number of primers were used for PCR amplification of Fd sequences. The library is expected to contain only clones expressing kappa/gamma 25 sequences. However, this is not an inherent limitation of the method since additional primers can be added to amplify any antibody class or subclass. Despite this restriction we were able to isolate a large number of antigen binding clones. Of interest is how a phage 30 library prepared as described herein compares with the in vivo antibody repertoire in terms of size, characteristics of diversity, and ease of access.

The size of the mammalian antibody repertoire is difficult to judge but a figure of the order of  $10^6$ - $10^8$  35 different antigen specificities is often quoted. With some of the reservations discussed below, a phage library of this size or large can readily be constructed by a



modification of the current method. In fact once an initial combinatorial library has been constructed, heavy and light chains can be shuffled to obtain libraries of exceptionally large numbers.

5 In principle, the diversity characteristics of the naive (unimmunized) in vivo repertoire and corresponding phage library are expected to be similar in that both involve a random combination of heavy and light chains. However, different factors will act to restrict the  
10 diversity expressed by an in vivo repertoire and phage library. For example a physiological modification such as tolerance will restrict the expression of certain antigenic specificities from the in vivo repertoire but these specificities may still appear in the phage library. For  
15 example, the representation of mRNA for sequences expressed by stimulated B-cells can be expected to predominate over those of unstimulated cells because of higher levels of expression. Different source tissues (e.g., peripheral blood, bone marrow or regional lymph  
20 nodes) and different PCR primers (e.g., ones expected to amplify different antibody classes) may result in library with different diversity characteristics.

Another difference between in vivo repertoire and phage library is that antibodies isolated from the former  
25 may have benefited from affinity maturation due to somatic mutations after combination of heavy and light chains whereas the latter randomly combines the matured heavy and light chains. Given a large enough phage library derived from a particular in vivo repertoire, the original matured  
30 heavy and light chains will be recombined. However, since one of the potential benefits of this new technology is to obviate the need for immunization by the generation of a single highly diverse "generic" phage library, it would be useful to have methods to optimize sequences to compensate  
35 for the absence of somatic mutation and clonal selection. Three procedures are made readily available through the methods of the present invention. First, saturation muta-

genesis may be performed on the CDR's and the resulting Fabs can be assayed for increased function. Second, a heavy or a light chain of a clone which binds antigen can be recombined with the entire light or heavy chain libraries respectively in a procedure identical to the one used to construct the combinatorial library. Third, iterative cycles of the two above procedures can be performed to further optimize the affinity or catalytic properties of the immunoglobulin. It should be noted that the latter two procedures are not permitted in B-cell clonal selection which suggests that the methods described here may actually increase the ability to identify optimal sequences.

Access is the third area where it is of interest to compare the in vivo antibody repertoire and phage library. In practical terms the phage library is much easier to access. The screening methods allow one to survey at least 50,000 clones per plate so that  $10^6$  antibodies can be readily examined in a day. This factor alone should encourage the replacement of hybridoma technology with the methods described here. The most powerful screening methods utilize selection which may be accomplished by incorporating selectable markers into the antigen such as leaving groups necessary for replication of auxotrophic bacterial strains or toxic substituents susceptible to catalytic inactivation. There are also further advantages related to the fact that the in vivo antibody repertoire can only be accessed via immunization which is a selection on the basis of binding affinity. The phage library is not similarly restricted. For example, the only general method to identify antibodies with catalytic properties has been by pre-selection on the basis of affinity of the antibody to a transition state analogue. No such restrictions apply to the in vivo library where catalysis can, in principle, be assayed directly. The ability to directly assay large numbers of antibodies for function may allow selection for catalysts in reactions where a mechanism is

**SUBSTITUTE SHEET**

not well defined or synthesis of the transition state analog is difficult. Assaying for catalysis directly eliminates the bias of the screening procedure for reaction mechanisms pejorative to a synthetic analog and therefore simultaneous exploration of multiple reaction pathways for a given chemical transformation are possible.

Although we have given examples of several screening methods, it should be clear to one skilled in the art that alternative methods of screening, such as by panning cells or particles expressing the protein product on their surface would essentially be equivalent. If the expressed gene products of interest are RNA molecules instead of proteins, screening could be accomplished by nucleic acid hybridization or by detecting some functional property of the RNA, such as ribozyme catalysis.

The methods disclosed herein describe generation of Fab fragments which are clearly different in a number of important respects from intact (whole) antibodies. There is undoubtedly a loss of affinity in having monovalent Fab antigen binders but this can be compensated by selection of suitably tight binders. For a number of applications such as diagnostics and biosensors it may be preferable to have monovalent Fab fragments. For applications requiring Fc effector functions, the technology already exists for extending the heavy chain gene and expressing the glycosylated whole antibody in mammalian cells.

The ideas presented here address the bottle neck in the identification and evaluation of antibodies. It is now possible to construct and screen at least three orders of magnitude more clones with mono-specificity than previously possible. The potential applications of the method should span basic research and applied sciences.

#### 21. Oligonucleotide Primer Design for Producing Dicistronic DNA

A method based on PCR amplification that fuses heavy and light chain sequences has been used to construct a

complete antigen binding domain of a Fab protein fragment composed of a heavy and a light chain. Schematic diagrams of an immunoglobulin molecule composed of heavy and light chains containing constant and variable regions is shown in Figure 1. Human heavy chain IgG and human kappa light chain are diagrammatically sketched in Figures 2A and 2B, respectively. To accomplish this procedure, immunoglobulin heavy and light chain primers were designed to produce a region of homology between two polymerase chain reaction (PCR) products. The complementary regions have been shown to hybridize predominantly under conditions where one set of primers ("inside primer pair") is used in a limiting amount relative to the other set of primers ("outside primer pair"). After the 3' ends of the PCR products have hybridized, the DNA polymerase has been shown to extend the ends creating a fusion sequence carrying the unique sequences of both PCR fragments separated by one copy of region X cistronic bridge. A two-step cloning procedure is thus avoided. When the recombinant sequence is then inserted into an expression vector such as ImmunoZAP, a fusion production capable of simultaneously expressing the heavy and light chains can be produced.

The strategy used for producing immunoglobulin heavy and light chain PCR dicistronic DNA is shown schematically in Figure 21. Regions of the immunoglobulin heavy chain coding strand are designated  $V_H$ ,  $C_H1$ ,  $C_H2$ , and  $C_H3$  corresponding to functional regions in the protein. The corresponding regions of the non-coding strand are designated by a prime ('). Regions  $V_L$  and  $C_L$  are similarly labelled for the kappa light chain. This procedure can also be performed using lambda light chain specific regions. A region, X, unrelated to the natural immunoglobulin sequences, is introduced into the fusion product by attaching X to the 5' ends of both of the  $C_H1'$  and  $V_L$  inside primers.

Overlapping oligonucleotide primers used in the fusion-PCR reactions to produce dicistronic DNA were

**SUBSTITUTE SHEET**

designed to encode the following: amino acids of 225 to 230 of the IgG heavy chain hinge region which are common to all human IgG isotypes; an Spe I restriction site; two stop codons; a ribosome binding site; a periplasmic (pelB) leader sequence (Better, et al., Science, 240:1041-1043 (1988); Lei, et al., J. Bacteriol., 169:4379-4383 (1988)); a Sac I restriction site which encodes amino acids 1 and 2 of the mature kappa light chain; and amino acids 3 to 8 of the mature kappa light chain. The X region was designed to contain a ribosome binding site and a pelB leader to ensure expression of the light chain. Nucleotide sequences for all human and mouse PCR primers, both inside and outside, are listed in Table 11. Primers followed by a prime (') represent non-coding strand sequences.

Table 11

## Human and Mouse PCR Primers

## Seq.

	<u>Id. No.</u>	<u>Human</u>	
20	(117)	V <sub>H</sub>	5'-GTCCTGTCCGAGGTGCAGCTGCTCGAGTCTGG-3'
	(118)	C <sub>H</sub> 1'	5'-AATAACAATCCAGCGGCTGCCGTAGGCAATAGGT ATTTTCATTATGACTGTCTCCTTGCTATTAAGT TACAAGATTTGGGCTC-3'
	(119)	V <sub>L</sub>	5'-GCCTACGGCAGCCGCTGGATTGTTATTAATCGCT GCCCCAACCTGCCATGGCTGAGCTCGTGATGACCC CAGTCTCC-3'
25	(120)	C <sub>L</sub> '	5'-TCCTTCTAGATTACTAACACTCTCCCCTGTTGAA GCTCTTTGTGACGGGCGAACTC-3'
		<u>Mouse</u>	
30	(121)	V <sub>H</sub>	5'-AGGTCCAGCTGCTCGAGTCTGG-3'
	(122)	C <sub>H</sub> 1'	5'-AATAACAATCCAGCGGCTGCCGTAGGCAATAGG TATTTTCATTATGACTGTCTCCTTGCTATTAAGT AGTATACAATCCCTGGGCACAAT-3'
	(123)	V <sub>L</sub>	5'-GCCTACGGCAGCCGCTGGATTGTTATTAATCGC TGCCCCAACCTGCCATGGCTGAGCTCGTGATGAC CCAGTCTCC-3'
35			

**SUBSTITUTE SHEET**

(124) C<sub>L</sub>' 5'-TCCTTCTAGATTACTAACACTCTCCCCTGTTGAA-3'

The overlapping regions of the human C<sub>H</sub>1' inside and V<sub>L</sub> inside primers are illustrated in Figure 22. The heavy chain downstream C<sub>H</sub>1' inside primer sequence is written 3' to 5' and the light chain upstream V<sub>L</sub> inside primer sequence is written 5' to 3'. The complementary PCR product strands, and not the primer strands, cross-prime to create the dicistronic molecule. Bold nucleotides represent regions where the C<sub>H</sub>1' inside primer hybridizes to the 3' end of C<sub>H</sub>1 on human IgG heavy chain mRNA or where the V<sub>L</sub> inside primer hybridizes to the 5' end of V<sub>L</sub> framework on human kappa light chain cDNA. The amino acid and nucleotides in italics represent changes in sequence from the original pelB leader sequence.

At amino acid 15 of the pelB leader sequence, the codon was changed from CTC to ATC resulting in a conservative amino acid change from a leucine to an isoleucine as shown in Figure 22 and Table 11. Hydrophobic amino acids in the core region of periplasmic leader sequences have been shown to be essential for correct processing of the leader sequence and transport of the mature protein to the periplasm. Oliver, in Neidhardt, R.C. (ed.), Escherichia coli and Salmonella Typhimurium, Am. Soc. Microbiol., 1:56-69 (1987). The nucleotide changes were made to allow for the artifactual insertion of one or two dATPs at the 3' end of the overlapping dicistronic molecules. Thermus aquaticus (Taq) DNA polymerase may add a dATP to the 3' end of the PCR product because of terminal transferase activity. Jiang, etg al. Oncogene, 4:923-928 (1989). The additional dATP would then cause a mismatch between the overlapping PCR products at the 3' terminus and inhibit elongation by Taq DNA polymerase. Sommer, et al. Nucl. Acids Res., 17:6749 (1989). Therefore, the change to two dTTPs in this position of the oligonucleotide primers would allow proper base pairing if up to two dATPs were added to the 3' terminus of the heavy chain PCR product.

The kappa light chain PCR product was designed to terminate at a position where two dTTPs occur 5' of the end of the product and did not require alterations of the nucleotide sequence. Nucleotides were changed in the kappa  
5 light chain primer encoding the pelB leader sequence without introducing amino acid changes in order to decrease the number of mismatches between the primer and the leader sequence of the kappa light chain mRNA as shown in Figure 22 and Table 11.

10 All primers were synthesized on an Applied Biosystems DNA synthesizer, Model 381A, following the manufacturer's instructions.

22. Preparation of a  $V_H$ -and  $V_L$ -Coding Repertoire

15 A. Preparation of a  $V_H$ -and  $V_L$ -Coding REpertoire from a Human cDNA Combinatorial Library

Cloned DNA, previously isolated from a combinatorial library that encodes human Fab fragments which bind tetanus toxoid (TT) was used as a template for preparing a  $V_H$ -and  $V_L$ -coding repertoire. Mullinax, et al., *supra*.  
20 Briefly, the combinatorial library was prepared by the following approach. Volunteer donors, who had been previously immunized against tetanus but had not received booster injections within the last year, received injections on 2 consecutive days of 0.5 milliliters (ml) of  
25 alum-absorbed tetnus toxoid (TT) (40 microgram/ml (ug)/ml) (Connaught Laboratories, Swiftwater, Pennsylvania).

One hundred ml of blood was drawn from the volunteers 6 days post injection and anticoagulated with a mixture of 0.14 M citric acid, 0.2 M trisodium citrate, and 0.22 M  
30 dextrose. The peripheral blood lymphocytes (PBLs) were recovered and isolated from the whole blood by layering the whole blood on Histopaque-1077 (Sigma, St. Louis, Missouri) and centrifuging at 400 x g for 30 minutes at 25 degrees Celsius (25°C). Isolated PBLs were washed twice  
35 with phosphate buffered saline (PBS) (150 mM sodium chloride and 150 mM sodium phosphate, pH 7.2 at 25°C).

Total RNA was then purified from the PBLs ( $10^6$  B cells per ml blood per 100 ml of blood) for an enriched source of B-cell mRNA coding for antiTT IgG using an RNA isolation kit according to manufacturer's instructions (Stratagene, La Jolla, California) and also described by Chomczynski et al., Anal. Biochem., 162:156-159 (1987). Briefly, the isolated PBLs were homogenized in 10 ml of a denaturing solution containing 4.0 M guanine isothiocyanate, 0.25 M sodium citrate at pH 7.0, and 0.1 M beta-mercaptoethanol. One ml of sodium acetate at a concentration of 2 M at pH 4.0 was admixed with the homogenized cells. Ten ml of phenol that had been previously saturated with  $H_2O$  was also admixed to the denaturing solution containing the homogenized cells. Two ml of a chloroform: isoamyl alcohol (24:1 v/v) mixture was added to this homogenate. The homogenate was mixed vigorously for ten seconds and maintained on ice for 15 minutes. The homogenate was then transferred to a thick-walled 50 ml polypropylene centrifuged tube (Fisher Scientific Company, Pittsburgh, Pennsylvania). The solution was centrifuged at 10,000 x g for 20 minutes at 4°C. The upper RNA-containing aqueous layer was transferred to a fresh 50 ml polypropylene centrifuge tube and mixed with an equal volume of isopropyl alcohol. This solution was maintained at -20°C for at least one hour to precipitate the RNA. The solution containing the precipitated RNA was centrifuged at 10,000 x g for twenty minutes at 4°C. The pelleted total cellular RNA was collected and dissolved in 3 ml of the denaturing solution described above. Three ml of isopropyl alcohol was added to the re-suspended total cellular RNA and inverted to mix. This solution was maintained at -20°C for at least 1 hour to precipitate the RNA. The solution containing the precipitated RNA was centrifuged at 10,000 x g for ten minutes at 4°C. The pelleted RNA was washed once with a solution containing 75% ethanol. The pelleted RNA was dried under vacuum for



15 minutes and then re-suspended in diethyl pyrocarbonate (DEPC) treated (DEPC-H<sub>2</sub>O) H<sub>2</sub>O).

Messenger RNA (mRNA) was prepared from the total cellular RNA using methods described in Molecular Cloning  
5 A Laboratory Manual, Maniatis et al., eds., Cold Spring Harbor, NY, (1982). Briefly, 500 mg of the total RNA isolated from a PBLs prepared as described above was re-suspended in one ml of 1X sample buffer (1 mM Tris-HCl, (Tris [hydroxymethyl-aminomethane]) pH 7.5; 0.1 mM EDTA  
10 (disodium ethylene diamine tetra-acetic acid), 0.5 M NaCl) and maintained at 65°C for five minutes and then on ice for five more minutes. The mixture was then applied to an oligo-dT (Stratagene) column that was previously prepared by washing the oligo-dT with a solution containing 10 mM  
15 Tris-HCl, pH 7.5; 1 mM EDTA, 0.5 M NaCl. The eluate was collected in a sterile polypropylene tube and reapplied to the same column after heating the eluate for five minutes at 65°C. The oligo dT column was then washed with 0.4 ml of high salt loading buffer consisting of 10 mM Tris-HCl  
20 at pH 7.5, 500 mM sodium chloride, and 1 mM EDTA. The oligo dT column was then washed with 2 ml of 1 X low salt buffer consisting of 10 mM Tris-HCl at pH 7.5, 100 mM sodium chloride, and 1 mM EDTA. The messenger RNA was eluted from the oligo dT column with 0.6 ml of buffer  
25 consisting of 10 mM Tris-HCl at pH 7.5, and 1mM EDTA. The messenger RNA was purified by extracting this solution with phenol/chloroform followed by a single extraction with 100% chloroform. The messenger RNA was concentrated by ethanol precipitation and re-suspended in DEPC H<sub>2</sub>O.

30 The messenger RNA isolated by the above process contains a plurality of different V<sub>H</sub> and V<sub>L</sub> coding polynucleotides, i.e., greater than about 10<sup>4</sup> different V<sub>H</sub>- and V<sub>L</sub>-coding genes.

Isolated RNA was converted to cDNA by a primer extension  
35 sion reaction with a first-strand synthesis kit according to manufacturer's instructions (Stratagene) by using an oligo (dT) primer for the light chain and a specific

**SUBSTITUTE SHEET**

primer, C<sub>H</sub>1', for the heavy chain. Mullinax et al., supra. In a typical 50 µl transcription reaction, 5 ug of PBL mRNA in water was first hybridized (annealed) with 200 ng (50.0 pmol) of an oligo (dT) primer for the light chain.

5 In a separate reaction, 5 ug of PBL mRNA in water was first hybridized (annealed with 200 ng (20 pmol) of the heavy chain primer, C<sub>H</sub>1', at 65°C for five minutes. Subsequently, the mixture was adjusted to 0.5 mM each of dATP, dCTP, dGTP and dTTP, 50 mM Tris-HCl at pH 8.3, 3 mM  
10 MgCl<sub>2</sub>, 75 mM KCl, 10 mM DTT, 20 units of RNase block II (Stratagene), and 20 units of Moloney-Murine Leukemia virus reverse transcriptase (Stratagene Cloning Systems), was added and the solution was maintained for 1 hour at 37°C. PCR amplification of the heavy and light chain  
15 sequences was done separately using 0.25-0.5 ug of first-strand synthesis product as template with sets of primer pairs using Taq DNA polymerase as described in Example 23.

The PCR amplified light chain DNA fragments were then digested with Sac I and Xba I and ligated into a modified  
20 Lambda Zap II vector as prepared in Example 29 to form a light chain ImmunoZap Library (ImmunoZAP L; Stratagene, La Jolla, California). The PCR amplified heavy chain DNA was digested with Spe I and Sho I and ligated into a different modified Lambda Zap II vector as prepared in Example 27 to  
25 form a heavy chain ImmunoZap Library (ImmunoZAP H; Stratagene). The resulting libraries were amplified and the resulting DNA was packaged into bacteriophage with in vitro packaging extract, Gigapack II gold (Stratagene) and used to infect E. coli strain XL1-Blue (Stratagene).

30 To construct a library for coexpression, the right arm of the heavy chain library phage DNA was digested with Hind III, preserving the left arm of ImmunoZAP H with a heavy chain inserts. The left arm of the light chain library phage DNA was digested with Mlu I resulting in a  
35 right arm of ImmunoZAP with kappa light chain inserts. Both products were then digested with EcoRI and ligated to create a combinatorial library that encoded human Fab

fragments including those specific for TT. Mullinax, et al., supra.

Reactive plaques were first identified by binding to tetanus toxoid as described in Example 31. Bacteriophage  
5 from purified reactive plaques were then converted to the plasmid format by in vivo excision with R408 helper phage (Stratagene) following methods described in Example 31 and familiar to one skilled in the art. Short, et al., Nucl. Acids. Res., 16:7583-7600 (1988). The resulting purified  
10 plasmid DNA encoding heavy and light chain was then used in PCR reactions as described below in Example 23.

B. Preparation of a V<sub>H</sub>- and V<sub>L</sub>-Coding Repertoire from mRNA from Tissues and Cells

(i) Human

15 Purified populations of PBLs, other lymphocytes, and hybridomas which express immunoglobulins including IgG, IgM, IgE, IgD, and IgA are used as sources for isolating mRNA encoding immunoglobulins. PBL's and other immunoglobulin expressing lymphocytes are isolated from either  
20 spleen, lymphoid tissue or plasma. Following purification of the cells, total RNA is then purified from these cells using a RNA isolation kit (Stratagene) as described in Example 22a. The purified RNA is then converted to cDNA with a first-strand synthesis kit as described in Example  
25 22a. The resultant cDNA is then used as a template in PCR amplification reactions as described below in Example 23 for the production of dicistronic molecules expressing heavy and light chains.

(ii) Mouse

30 Populations of cells described above can be isolated from other mammalian sources such as mouse or rabbit. Both mRNA and rearranged DNA can be isolated as described above and used as templates in PCR amplification reactions. cDNA synthesized from mRNA isolated from a mouse  
35 anti-human fibronectin hybridoma (ATCC, CRL-1606) was used

as a preferred template for the production of dicistronic molecules expressing heavy and light chain.

c. Preparation of a  $V_H$ -Coding Repertoire from Rearranged DNA

5 Rearranged DNA isolated from PBLs, other lymphocytes, and hybridomas which express immunoglobulins can be used to prepare a  $V_H$ -coding repertoire. The amplification procedure for preparing a  $V_H$ -coding repertoire using rearranged DNA is performed as described in Example 23.

10 23. Preparation of DNA Homologs

A.  $V_H$ -Coding Double Stranded DNA Homologs

Cloned DNA, prepared in Example 22 from a combinatorial library that encodes human Fab fragments which bind tetanus toxoid (TT), was used as a template for preparing  
15 a  $V_H$ -coding double stranded DNA homolog. Human heavy chain, containing both the  $V_H$  and  $C_H1$  coding region and designated as Fd, was amplified in a PCR reaction. The amplification was performed in a 100 ul reaction containing 5 nanograms (ng) of the cloned DNA in PCR buffer  
20 consisting of the following: 10 mM Tris-HCl, pH 8.3; 50 mM KCl, 1.5 mM  $MgCl_2$ ; 0.001% (w/v) gelatin; 200 mM of each dNTP; 200 nanomolar (nM) of each primer; and 2.5 units of Taq DNA polymerase. The human  $V_H$  outside primer and  $C_H1$ ' inside primer were used as a PCR primer pair for amplification  
25 of the heavy chain (Table 11 and Figure 21). The reaction mixture was overlaid with mineral oil and subjected to 40 cycles of amplification. Each amplification cycle (thermocycle) involved denaturation at 94°C for 1.5 minutes, annealing at 54°C for 2.5 minutes and polynucleotide synthesis by primer extension (elongation) at  
30 72°C for 3.0 minutes followed by a return to the denaturation temperature. The resultant amplified  $V_H$ -coding DNA homolog containing samples were then gel purified, extracted twice with phenol/chloroform, once with chloro-

form followed by ethanol precipitation and were stored at -70°C in 10 mM Tris-HCl, pH 7.5, and 1 mM EDTA.

To verify the amplification of the heavy chain, the PCR purified products were electrophoresed in an agarose gel. The expected size of the heavy chain was approximately 730 base pairs as shown in Figure 23. The  $V_H$ -coding double stranded DNA homologs were then used in subsequent PCR amplification reactions with  $V_L$ -coding counterparts prepared below for the production of dicistronic DNA molecules having  $V_H$  and  $V_L$  cistronic portions as illustrated in Example 24.

#### B. $V_L$ -Coding Double Stranded DNA Homologs

Cloned DNA, prepared in Example 22 from a combinatorial library that encodes human Fab fragments which bind tetanus toxoid (TT), was used as a template for preparing a  $V_L$ -coding double stranded DNA homolog. Human light chain, containing the entire coding region of kappa light chain ( $V_L$  and  $C_L$ ), was amplified using the same PCR conditions described for human heavy chain with the exception that a human  $V_L$  inside primer and  $C_L'$  outside primer were used as the PCR primer pair (Table 11 and Figure 21). The resultant  $V_L$ -coding double stranded DNA homolog was gel purified and stored as described above.

To verify the amplification of the light chain, the PCR purified products were electrophoresed in an agarose gel. The expected size of the light chain was approximately 690 base pairs as shown in Figure 23. The  $V_L$ -coding double stranded DNA homologs were then used in subsequent PCR amplification reactions with  $V_H$ -coding counterparts prepared above for the production of dicistronic DNA molecules as illustrated in Example 24.

24. Preparation of Internally-Primed Duplexes of  $V_H$ - and  $V_L$ -Coding DNA Homolog

A. Hybridization of  $V_H$ - with  $V_L$ -Coding DNA Homologs

The  $V_H$ - and  $V_L$ -coding double stranded DNA homologs  
5 prepare in Examples 23A and 23B, respectively, were  
admixed together and denatured at 95°C for 5 minutes to  
separate the strands of each homolog. The denatured  $V_H$ -  
and  $V_L$ -coding DNA strands in the admixture were then  
annealed at 54°C for 5 minutes to form a  $V_H$ - and  $V_L$ -coding  
10 duplex DNA molecule hybridized at the 3' ends at region X  
of each original homolog. One strand of the X region  
(cistronic) bridge encodes at least one stop codon in the  
same reading frame as the upstream cistron, a ribosome  
binding site downstream from the stop codon, and a  
15 polypeptide leader (pelB) having a translation initiation  
codon in the same reading frame as the downstream cistron  
located downstream from the ribosome binding site.

B. Primer Extension to Produce Dicistronic DNA Molecules

The hybridized recombinant  $V_H$ - and  $V_L$ -coding DNA  
20 molecule (internally primed duplex) was subjected to  
primer extension and then amplified with only the  $V_H$  and  
 $C_L$ ' primers following the PCR reaction procedure described  
in Example 23A. This second PCR reaction is schematically  
represented in Figure 21. The PCR reaction products were  
25 gel electrophoresed to verify the presence of the result-  
ant  $V_H$ - and  $V_L$ -coding dicistronic DNA molecules. The  
expected size of the dicistronic molecule was about 1390  
base pairs and is shown in Figure 23. The resultant  $V_H$ -  
and  $V_L$ -coding dicistronic DNA molecules were then ligated  
30 into the modified ImmunoZAP H vector (Figures 24A and 24B)  
for the construction of expression vectors as described in  
Example 30.

2. Preparation of Mouse Hybridoma  $V_H$ - and  $V_L$ -Coding Double Stranded DNA Homologs and Production of Dicistronic DNA Molecules in a Single Amplification Reaction

5 Mouse hybridoma heavy and light chain cDNA prepared in Example 22B was amplified in a single PCR reaction using the reaction conditions given above with an excess of the outside primers (200 nM concentration of both the mouse  $V_H$  primer and  $C_L'$  primer) and a limiting amount of  
10 the inside primers (20 nM concentration of both the mouse  $C_H1'$  and  $V_L$  primer) (Table 11). The resultant mouse heavy and light chain dicistronic molecules were then inserted into a modified ImmunoZAP H for construction of an expression vector as described in Example 30.

15 26. Preparation of Internally-Primed Duplexes Using a Single Internal Primer that Overlaps Both the  $V_H$  and  $V_L$  Repertoires

Another approach to producing a library of dicistronic DNA molecules is to use a single internal primer  
20 instead of using two separately internal primers. The process of creating a dicistronic molecule comprising an upstream  $V_H$  cistron and a downstream  $V_L$  cistron is to combine in a PCR buffer the following: a repertoire of  $V_H$  genes consisting of at least  $10^5$  different genes; a reper-  
25 toire of  $V_L$  genes consisting of at least  $10^4$  different genes; an outside  $V_H$  primer; an outside  $V_L$ ; and a polynucleotide strand having a 3'-terminal priming portion, a cistronic bridge coding portion, and a 5' terminal primer-template portion. The PCR reaction is performed as  
30 described in Example 22A.

The 3'-terminal priming portion of a polynucleotide strand (linker) has a nucleotide base sequence homologous to a portion of the primer extension product of one of the outside primers. The 5'-terminal priming portion encodes  
35 a nucleotide base sequence homologous to a portion of the primer extension product of the other outside primer. The

cistronic bridge coding portion encodes at least one stop codon in the same reading frame as the upstream cistron, a ribosome binding site downstream from the stop codon and a polypeptide leader (pelB) having a translation initiation codon in the same reading frame as the downstream cistron where the initiation codon is located downstream from the ribosome binding site. Polynucleotide strand (linker) primers useful in this invention are listed in Table 12.

10 Table 12

Polynucleotide Strand (Linker) Primers

Seq.

Id. No.

	(1251) <sup>1</sup>	1'	5' GGAGAGTGGGTCATCAGAGCTCAGCCATGGCAGGTTGG GCAGCGATTAATAACAATCCAGCGGCTGCCGTAGGCAAT AGGTATTTTCATTATGACTGTCTCCTTGCTATTAAGTAGT ACAAGATTTGGGCTC 3'
15			
	(126) <sup>2</sup>	2'	5' GAGCCCAAATCTTGTACTAGTTAATAGCAAGGAGACAGT CATAATGAAATACCTATTGCCTACGGCAGCCGCTGGATT GTTATTAATCGCTGCCCAACCTGCCATGGCTGAGCTCGT GATGACCCACTCTCC 3'
20			

<sup>1</sup> Primes mRNA (sense strand) of heavy chain C<sub>H</sub>1 region; antisense strand of light chain V<sub>L</sub> with dicistronic bridge in between heavy and light chains will be in the same relative orientation as given in the example.

<sup>2</sup> Primes antisense strand of heavy chain C<sub>H</sub>1 regions; and sense strand of light chain V<sub>L</sub> region with dicistronic in between heavy and light chains will be in the same relative orientation as given in the example.

30 The resultant single step internally primed dicistronic DNA molecule can then be ligated into modified ImmunoZAP H for construction of an expression vector as described in Example 30.



## 27. Preparation of Lambda Zap II Expression Vector

The vector Lambda Zap<sup>TM</sup> II (Stratagene) is a derivative of the original Lambda Zap (ATCC # 40,298) that maintains all of the characteristics of the original Lambda Zap including 6 unique cloning sites, fusion protein expression, and the ability to rapidly excise the insert in the form of a phagemid (Bluscript SK-), but lacks the SAM 100 mutation, allowing growth on many Non-Sup F strains, including XL1-Blue. The Lambda Zap II was constructed as described in Short et al., Nucleic Acids Res., 16:7583-7600, (1988), by replacing the Lambda S gene contained in a 4254 base pair (bp) DNA fragment produced by digesting Lambda Zap with the restriction enzyme NcoI. This 4254 bp DNA fragment was replaced with the 4254 bp DNA fragment containing the Lambda S gene isolated from Lambda gt10 (ATCC # 40,179) after digesting the vector with the restriction enzyme NcoI. The 4254 bp DNA fragment isolated from lambda gt10 was ligated into the original Lambda Zap vector using T4 DNA ligase and standard protocols for such procedures described in Current Protocols in Molecular Biology, Ausubel et al., eds., John Wiley and Sons, NY, 1987.

## 28. Preparation of V<sub>H</sub>-Expression Vectors, ImmunoZAP H and Modified ImmunoZAP H, Construction

### 25 A. ImmunoZAP H

The main criterion used in choosing a vector system was the necessity of generating the largest number of Fab fragments which could be screened directly. Bacteriophage lambda was selected as the expression vector for three reasons. First, in vitro packaging of phage DNA is the most efficient method of reintroducing DNA into host cells. Second, it is possible to detect protein expression at the level of single phage plaques. Finally, the screening of phage libraries typically involve less difficulty with nonspecific binding. The alternative, plasmid cloning vectors, are only advantageous in the analysis of

clones after they have been identified. This advantage is not lost in the present system because of the use of lambda Zap, thereby permitting a plasmid containing the heavy chain, light chain, or Fab expressing inserts to be  
5 excised.

To express the plurality of  $V_H$ -coding DNA homologs in an *E. coli* host cell, a vector was constructed that placed the  $V_H$ -coding DNA homologs in the proper reading frame, provided a ribosome binding site as described by Shine et  
10 al., Nature, 254:34, (1975), provided a leader sequence directing the expressed protein to the periplasmic space, provided a polynucleotide sequence that coded for a known epitope (epitope tag) and also provided a polynucleotide that coded for a spacer protein between the  $V_H$ -coding DNA  
15 homolog and the polynucleotide coding for the epitope tag. A synthetic DNA sequence containing all of the above polynucleotides and features was constructed by designing single stranded polynucleotide segments of 20-40 bases that would hybridize to each other and form the double  
20 stranded synthetic DNA sequence shown in Figure 25A. The individual single-stranded polynucleotides ( $N_1$ - $N_{12}$ ) are shown in Table 13 below.

Table 13

Seq.

25 Id. No.

- (91) N1) 5' GGCCGCAAATTCTATTTCAAGGAGACAGTCAT 3'  
 (92) N2) 5' AATGAAATACCTATTGCCTACGGCAGCCGCTGGATT 3'  
 (93) N3) 5' GTTATTACTCGCTGCCCCAACCAGCCATGGCCC 3'  
 (94) N4) 5' AGGTGAAACTGCTCGAGAATTCTAGACTAGGTTAATAG 3'  
 30 (95) N5) 5' TCGACTATTAAGTAGTCTAGAATTCTCGAG 3'  
 (96) N6) 5' CAGTTTCACCTGGGCCATGGCTGGTTGGG 3'  
 (97) N7) 5' CAGCGAGTAATAACAATCCAGCGGCTGCCGTAGGCAATAG 3'  
 (98) N8) 5' GTATTTCAATTATGACTGTCTCCTTGAAATAGAATTTGC 3'  
 (99) N9-4) 5' AGGTGAAACTGCTCGAGATTTCTAGACTAGTTACCCGTAC 3'  
 35 (100) N11) 5' GACGTTCCGGACTACGGTTCTTAATAGAATTCG 3'  
 (101) N12) 5' TCGACGAATTCTATTAAGAACCGTAGTC 3'

(102) N10-5) 5' CGGAACGTCGTACGGGTAAGTCTAGAAATCTCGAG 3'

Polynucleotide N2, N3, N9-4', N11, N10-5', N6, N7 and N8 were kinased by adding 1  $\mu$ l of each polynucleotide (0.1 ug/ $\mu$ l) and 20 units of T<sub>4</sub> polynucleotide kinase to a solution containing 70 mM Tris-HCl at pH 7.6, 10 mM MgCl<sub>2</sub>, 5 mM DTT, 10 mM beta mercaptoethanol, 500 ug/ml of BSA. The solution was maintained at 37°C for 30 minutes and the reaction stopped by maintaining the solution at 65°C for 10 minutes. The two end polynucleotides, 20 ng, of polynucleotides N1 and polynucleotides N12, were added to the above kinasing reaction solution together with 1/10 volume of a solution containing 20 mM Tris-HCl, pH 7.4, 2 mM MgCl<sub>2</sub> and 50 mM NaCl. This solution was heated to 70°C for 5 minutes and allowed to cool to room temperature, approximately 25°C, over 1.5 hours in a 500 ml beaker of water. During this time period all 10 polynucleotides annealed to form the double stranded synthetic DNA insert shown in Figure 25A. The individual polynucleotides were covalently linked to each other to stabilize the synthetic DNA insert by adding 40  $\mu$ l of the above reaction to a solution containing 50 mM Tris-HCl, pH 7.5, 7 mM MgCl<sub>2</sub>, 1 mM DTT, 1 mM ATP and 10 units of T<sub>4</sub> DNA ligase. This solution was maintained at 37°C for 30 minutes and then the T<sub>4</sub> DNA ligase was inactivated by maintaining the solution at 65°C for 10 minutes. The end polynucleotides were kinased by mixing 52  $\mu$ l of the above reaction, 4  $\mu$ l of a solution containing 10 mM ATP and 5 units of T<sub>4</sub> polynucleotide kinase. This solution was maintained at 37°C for 30 minutes and then the T<sub>4</sub> polynucleotide kinase was inactivated by maintaining the solution at 65°C for 10 minutes.

The completed synthetic DNA insert was ligated directly into a lambda Zap II vector prepared in Example 27 that had been previously digested with the restriction enzymes NotI and XhoI. The ligation mixture was packaged according to the manufacturer's instructions using Gigapack II Gold packing extract (Stratagene). The pack-

aged ligation mixture was plated on XL1-blue cells (Stratagene). Individual Lambda Zap II plaques were cored and the inserted excised according to the in vivo excision protocol provided by the manufacturer (Stratagene). This  
 5 in vivo excision protocol converts the cloned insert from the Lambda Zap II vector into a plasmid vector to allow easy manipulation and sequencing. The accuracy of the above cloning steps was confirmed by sequencing the insert using the Sanger dideoxy method described in by Sanger et  
 10 al., Proc. Natl. Acad. Sci. USA, 74:5463-5467, (1977) and using the manufacturer's instructions in the AMV Reverse Transcriptase <sup>35</sup>S-ATP sequencing kit (Stratagene). The sequence of the resulting V<sub>H</sub> expression vector is shown in Figure 25A and Figure 26.

#### 15 B. Modified ImmunoZAP H

To create a fusion-PCR library from hybridoma RNA for expressing the plurality of V<sub>H</sub>-coding DNA homologs in an E. coli host cell, a vector based on the ImmunoZAP H vector described above was constructed. The procedure for con-  
 20 structing the vector was performed as described above with the following modifications: elimination of the SacI site between the T<sub>3</sub> polymerase and NotI sites and changing the nucleotide base residue sequence from AAA to CAG which resulted in an amino acid residue change from lysine to  
 25 glutamine as shown in Figures 24A and 24B.

The individual single-stranded polynucleotides (N<sub>1</sub>, N<sub>4</sub>, N<sub>6</sub> and N<sub>7</sub>), which were modified from their counterparts listed in Table 14, are listed in Table 14 below.

Table 14

30 Seq.

Id. No.

- (127) N1) 5' AGCTGCGGCCGCAAATTCTATTTCAAGGAGACAGTCAT 3'  
 (128) N2) 5' AATGAAATACCTATTGCCTACGGCAGCCGCTGGATT 3'  
 (129) N3) 5' GTTATTACTCGCTGCCCAACCAGCCATGGCCC 3'  
 35 (130) N4) 5' AGGTGCAGCTGCTCGAGAATTCTAGACTAGGTTAATAG 3'

(131) N5) 5' TCGACTATTAAGTCTAGAAATTCCTCGAG 3'  
 (132) N6) 5' CAGCTGCACCTGGGCCATGGCTGGTTGGG 3'  
 (133) N7) 5' CAGCGAGTAATAACAATCCAGCGGCTGCCGTAGGCAATAG 3'  
 (134) N8) 5' CTATTTTCATTATGACTGTCTCCTTGAAATAGAATTTGCGGCCGC  
 5 3'  
 (135) N9-4) 5' AGGTGAAACTGCTCGAGATTTCTAGACTAGTTACCCGTAC 3'  
 (136) N11) 5' GACGTTCCGGACTACGGTTCCTTAATAGAATTCG 3'  
 (137) N12) 5' TCGACGAATTCCTATTAAGAACCGTAGTC 3'  
 (138) N10-5) 5' CGGAACGTCGTACGGGTAAGTCTAGAAATCTCGAG 3'

10 The modified ImmunoZAP H vector was created to eliminate an unnecessary SacI site in the ImmunoZAP H vector, (Example 28a, when the heavy and light chain vectors were combined. The modifications also improved the efficiency of secretion of positively charged amino  
 15 acids in the amino terminus of the expressed protein. Inouye et al., Proc. Natl. Acad. Sci. USA, 85:7685-7689 (1988).

## 29. Preparation of V<sub>L</sub> Expression Vector ImmunoZAP L Construction

20 To express the plurality of V<sub>L</sub> coding polynucleotides in an E. coli host cell, a vector was constructed that placed the V<sub>L</sub> coding polynucleotide in the proper reading frame, provided a ribosome binding site as described by Shine et al., Nature, 254:34, (1975), provided a leader  
 25 sequence directing the expressed protein to the periplasmic space and also provided a polynucleotide that coded for a spacer protein between the V<sub>L</sub> polynucleotide. A synthetic DNA sequence containing all of the above polynucleotides and features was constructed by designing  
 30 single stranded polynucleotide segments of 20-40 bases that would hybridize to each other and form the double stranded synthetic DNA sequence shown in Figure 25B. The individual single-stranded polynucleotides (N<sub>1</sub>-N<sub>8</sub>) are shown in Table 13 above.

35 Polynucleotides N2, N3, N4, N6, N7 and N8 were kinased by adding 1  $\mu$ l of each polynucleotide and 20 units

of T<sub>4</sub> polynucleotide kinase to a solution containing 70 mM Tris-HCl, pH 7.6, 10 mM MgCl<sub>2</sub>, 5 mM DDT, 10 mM 2ME, 500 micrograms per ml of BSA. The solution was maintained at 37°C for 30 minutes and the reaction stopped by maintaining the solution at 65°C for 10 minutes. The two end polynucleotides 20 ng of polynucleotides N1 and polynucleotides N5 were added to the above kinasing reaction solution together with 1/10 volume of a solution containing 20 mM Tris-HCl, pH 7.4, 2 mM MgCl<sub>2</sub> and 50 mM NaCl. This solution was heated to 70°C for 5 minutes and allowed to cool to room temperature, approximately 25°C, over 1.5 hours in a 500 ml beaker of water. During this time period all of the polynucleotides annealed to form the double stranded synthetic DNA insert. The individual polynucleotides were covalently linked to each other to stabilize the synthetic DNA insert with adding 40 µl of the above reaction to a solution containing 50 µl Tris-HCl, pH 7.5, 7 mM MgCl<sub>2</sub>, 1 mM DTT, 1 mM ATP and 10 units of T4 DNA ligase. This solution was maintained at 37°C for 30 minutes and then the T4 DNA ligase was inactivated by maintaining the solution at 65°C for 10 minutes. The end polynucleotides were kinased by mixing 52 µl of the above reaction, 4 µl of a solution recontaining 10 mM ATP and 5 units of T4 polynucleotide kinase. This solution was maintained at 37°C for 30 minutes and then the T4 polynucleotide kinase was inactivated by maintaining the solution at 65°C for 10 minutes.

The completed synthetic DNA insert was ligated directly into a Lambda Zap II vector prepared in Example 27 that had been previously digested with the restriction enzymes NotI and XhoI. The ligation mixture was packaged according to the manufacturer's instructions using Gigapack II Gold packing extract and the packaged ligation mixture was plated on XL1-Blue cells as described in Example 28A. Individual lambda Zap II plaques were cored and the inserts excised according to the in vivo excision protocol as described in Example 28A. This in vivo

excision protocol converts the cloned insert from the Lambda Zap II vector into a phagemid vector to allow easy manipulation and sequencing and also produces the phagemid version of the  $V_L$  expression vectors. The accuracy of the  
5 above cloning steps was confirmed by sequencing the insert using the Sanger dideoxy method described by Sanger et al., Proc. Natl. Acad. Sci. USA, 74:5463-5467, (1977) and using the manufacturer's instructions in the AMV reverse transcriptase  $^{35}\text{S}$ -dATP sequencing kit (Stratagene). The  
10 sequence of the resultant  $V_L$  expression vector is shown in Figure 25B and Figure 27).

The  $V_L$  expression vector used to construct the  $V_L$  library was the phagemid produced to allow the DNA of the  $V_L$  expression vector to be determined. The phagemid was  
15 produced, as detailed above, by the in vivo excision process from the Lambda Zap  $V_L$  expression vector (Figure 27).

### 30. Construction of $V_{HL}$ -Expression Vectors and Library

#### A. Ligation of Dicistronic DNA Molecules with Modified ImmunoZAP H

20 In preparation for cloning a library enriched in  $V_H$ - $V_L$ -coding ( $V_{HL}$ ) dicistronic DNA molecules, PCR amplified products (human or mouse) prepared in Examples 24, 25, and 26 (50 mM NaCl, 25 mM Tris-HCl, pH 7.7, 10 mM  $\text{MgCl}_2$ , 10 mM  $\beta$ -mercaptoethanol, 100 ug/ml BSA, at 37°C were digested  
25 with restriction enzymes XhoI and XbaI at a concentration of 60 units of enzyme per ug of DNA, and purified on a 1% agarose gel. After gel electrophoresis of the digested PCR amplified dicistronic DNA molecules, the region of the gel containing the DNA fragments of approximately 1360  
30 base pairs in size was excised, purified using Gene-Clean (BIO 101, La Jolla, California), ethanol precipitated and resuspended in 10 mM Tris-HCl, pH 7.5, and 1 mM EDTA to a final concentration of 10 ng/ul. Equimolar amounts of the insert were then ligated overnight at 4°C to 1 ug of  
35 modified ImmunoZAP H vector, prepared in Example 28b, (Stratagene) previously digested with XhoI and XbaI. A

portion of the ligation mixture (1 ul) was packaged for 2 hours at room temperature using Gigapack Gold packaging extract (Stratagene) and the packaged material was plated on a permissive E. coli (strain XL1-blue) lawn to generate  
5 plaques. The library was determined to consist of predominantly  $V_{HL}$  with less than 5% non-recombinant background.

## B. Screening of Antibody-Producing Plaques

### (i) Human

To screen for expression of  $V_{HL}$  dicistronic molecules,  
10 E. coli were infected to yield approximately 100 plaques per plate. Replica filter lifts of the plaques on an agar plate were produced by overlaying a nitrocellulose filter that had been soaked in 10 mM isopropyl beta-dithiogalactopyranoside on each plate with transfer for  
15 15 hours at 23°C. For detection of  $V_{HL}$  antibody fragment expression, the filters were screened with rabbit anti-human heavy and light chain antibodies followed by goat anti-rabbit antibody coupled to alkaline phosphatase (Cappel Laboratories, Malver, Pennsylvania). The detec-  
20 tion of immunoreactive product confirmed the presence and expression of  $V_{HL}$  antibody fragments.

To identify human DNA clones expressing antibody that bound TT, plaques were plated and proteins expressed as described above. Replica filters were incubated with 0.2  
25 nN  $^{125}$ I-tetanus toxoid and washed. Positive plaques were identified by autoradiography and isolated. The frequency of positive clones in the library was equivalent to (number of positive clones)/[number of plaques screened] X (fraction of plaques expressing  $V_{HL}$ ). Concentrated non-adsorbed tetanus toxoid was iodinated with sodium iodide  
30  $^{125}$ I (ICN, Irvine, California) by the Choramine-T method as described in Botton et al., Biochem. J., 133:529-539 (1973) and available in a kit (Iodo-Beads, Pierce, Rockford, Illinois).

35 Human DNA clones were re-plated at approximately 100 phage per plaque side by side with the parental phage that



were used as templates for PCR amplification and screened in the primary antigen binding screen. The results of the screening procedure are seen in Figure 28. Similar signals between the parental clones and the V<sub>HL</sub> dicistronic DNA molecules demonstrated that the sequence differences introduced with the C<sub>H</sub>1' and V<sub>L</sub> primers did not adversely affect gene expression. Also, it should be noted in Figure 28 that a random parental clone that did not react with tetanus toxoid, 7G1, was unreactive before and after the PCR dicistronic fusion, as was the control ImmunoZAP H vector (IZ H).

(ii) Mouse

Mouse antibody-producing plaques prepared in Example 27 were screened for antibody expression with rabbit anti-mouse heavy and light chain antibody (Cappel Laboratories) as described above.

31. Characterization of Cloned Dicistronic V<sub>HL</sub> Repertoire in Expression Library

A. Verification of Presence and Size of Cloned Dicistronic V<sub>HL</sub> Repertoire

Bacteriophage from purified reactive plaques prepared in Example 30B were converted to the plasmid format by in vivo excision with R408 helper phage according to manufacturer's protocol (Stratagene) and also described in Short et al., Nucl. Acids Res., 16:7583-7600 (1988). In the in vivo excision protocol, the cloned insert from the ImmunoZAP H vector was converted into a phagemid vector to allow easy manipulation and sequencing. Briefly, phage plaques were cored from the agar plates and transferred to sterile microfuge tubes containing 500 ul of a buffer containing 50 mM Tris-HCl, pH 7.5, 100 mM NaCl, 10 mM MgSO<sub>4</sub>, and 0.01% (w/v) gelatin and 20 ul of chloroform.

For excisions, 200 ul of the phage stock, 200 ul of XL1-Blue cells (A<sub>600</sub> = 1.00) and 1 ul of R408 helper phage (1 x 10<sup>10</sup> plaque forming units Opfu/ml) were incubated at

37°C for 15 minutes. After a 4 hour incubation in Luria-Bertani (LB) broth and heating at 70°C for 20 minutes to heat kill the XL1-blue cells, the phagemids were re-infected into XL1-Blue cells and plated onto LB plates  
5 containing ampicillin. Double stranded DNA was prepared from the phagemid containing cells according to the methods described by Holmes et al., Anal. Biochem., 114:193, (1981). Clones were first screened for DNA inserts by restriction digests with XhoI and XbaI. The detection of  
10 1390 base pair fragment on an agarose gel confirmed the presence of a  $V_{HL}$  dicistronic molecule insert.

B. Sequencing of Plasmids from Expression Library

Clones containing the putative  $V_{HL}$  insert were sequenced using reverse transcriptase according to the  
15 general method described by Sanger et al., Proc. Natl. Acad. Sci., USA, 74:5463-5467, (1977) and the specific modifications of this method provided in the manufacturer's instructions in the AMV reverse transcriptase  $^{35}\text{S}$ -dATP sequencing kit (Stratagene).

20 Nucleotide sequence analysis of several fusion clones indicated that the sequence of the fusion region was identical to that shown in Figure 22, proving that the clones were actually generated through a fusion PCR intermediate.

25 C. Advantages of Fusion-PCR to Produce Dicistronic DNA Molecules

PCR amplification can, therefore, be used to fuse sequences responsible for encoding subunits of a heterodimeric protein together into a single DNA fragment that  
30 can then direct the expression of both subunits from one expression vector. In the case of antibodies, if the source of nucleic acid template comes from hybridoma mRNA, there is only one heavy and light chain sequence to choose from, and thus the heavy:light pair is a "natural" pair.

However, if spleen, peripheral blood B-cell, or other lymphocyte mRNA is used as the source of template, the PCR fusion reaction to form a dicistronic DNA molecule can randomly pair heavy and light chains from different cells, producing a combinatorial library. In such a library, only a small fraction of the clones contain the original heavy and light chain pairs. This may not be a problem if the desired natural pair is well represented in the original B-cell population, as is the case with hyperimmunized donors. However, if one wishes to find a naturally occurring rare specificity in a combinatorial library, one may have to screen a large number of clones.

The fusion method presented here may offer a solution to the random combinatorial problem. If one begins with a very dilute population of B-cells (possibly in a medium that limits diffusion), it may be possible for the dicistronic event to occur between naturally paired heavy and light chain sequences before significant mixing between B-cell RNA occurs. Thus, the fused heavy and light chain sequences would be the original pairs, and the resulting library would express predominantly the naturally occurring antibody specificities. Such a library would be highly preferable when rare natural specificities are sought.

Another advantage to this method is that only one vector and one cloning step are necessary. This saves a substantial amount of time, resources, and effort. Moreover, the ease of the single PCR reaction greatly simplified the process of going from B-cell RNA to an *E. coli* library, making this approach a noteworthy alternative to standard hybridoma technology.

The foregoing is intended as illustrative of the present invention but not limiting. Numerous variations and modifications can be effected without departing from the true spirit and scope of the invention.

Claims:

1. A method of producing a nucleic acid vector encoding two or more desired genes, each from a family of genes, said genes being capable of together producing a characteristic that can be used to identify the vector encoding said desired genes from other vectors encoding other combinations of genes from said families of genes, which method comprises:

10 a) randomly inserting into vectors one member from a first family of genes and one member from one or more other families of genes so that a population of vectors are created wherein each vector may contain one of the genes from said first gene family and one of the genes from each of said other gene families;

15 b) identifying within said population of vectors a vector capable of detectably producing a desired characteristic resulting from the inclusion of one gene from said first gene family and one gene from each of said other gene families, and using said characteristic to  
20 distinguish the vector from other vectors within the population containing undesired combinations of gene members from said gene families.

2. The method of claim 1 wherein said genes are inserted into a DNA vector at one or more integration  
25 sites, which method further comprises:

a) preparing said vectors with one or more site- or region-specific recombination sequences;

b) permitting, in the presence of one or more reagents facilitating said site- or region-specific  
30 recombination, a member of said first family of genes to combine in a vector with a member of said second family of genes.

3. The method of claim 2 wherein said site- or region-specific recombination site is recognized and acted  
35 on by flp recombinase.

4. The method of claim 2 wherein said site- or region-specific recombination site is recognized and acted on by cre recombinase.

5 5. The method of claim 2 wherein said site- or region-specific recombination site is recognized and acted on by lambda integrase recombinase.

6. The method of claim 2 wherein at least one of the vectors contains a sequence capable of being recognized and acted on by transposase.

10 7. The method in claim 1 where said genes are inserted into a DNA vector at one or more integration sites, which method further comprises:

a) cleaving said vector with one or more site-specific integration reagents;

15 b) preparing the ends of genes from said first family of genes so that one end will ligate with an end of the vector cleaved by a first reagent and the other with an end of the vector cleaved by a second reagent;

20 c) preparing the ends of said genes from said other gene families so that one end will ligate with an end of the vector cleaved by a third reagent and the other with an end of the vector cleaved by a fourth reagent;

d) preparing at least one double stranded DNA linker fragment having one end ligatable to one end of  
25 said genes from said first family of genes and the other end ligatable to one end of genes from said other family of genes;

e) mixing said vector, genes, and said linker fragment or fragments together in a ligation mix and  
30 ligating the components.

8. The method of claim 7 wherein said reagents are the same.

**SUBSTITUTE SHEET**

9. the method of claim 8 wherein said reagents are different.

10. The method of claim 1, wherein said combination of genes is accomplished in vivo.

5        11. A method of producing a host cell expressing two or more desired genes, each from a family of genes, said genes being capable of together producing a characteristic that can be used to identify the host cell expressing said  
10 combinations of genes from said families of genes, which method comprises:

      a) randomly introducing into host cells one member from a first family of genes and one member from one or more other families of genes so that a population of host  
15 cells are created wherein each host cell may contain one of the genes from said first gene family and one of the genes from each of said other gene families;

      b) identifying within said population of host cells a host cell capable of detectably exhibiting a desired  
20 characteristic resulting from the inclusion of one gene from said first gene family and one gene from each of said other gene families, and using said characteristic to distinguish the host cell from other host cells within the population containing undesired combinations of gene  
25 members from said gene families.

12. The method of claim 11 wherein said vectors are lambda bacteriophage vectors and the host cells are E. Coli.

13. A method of producing a nucleic acid vector  
30 encoding two or more genes belonging to families of genes, being capable of producing a characteristic that can be used to identify the vector encoding said genes from other

vectors encoding other members of the families of genes which method comprises:

a) isolating a first population of vectors for which each member of said population may contain one  
5 member of a family of genes;

b) inserting one member of a second family of genes into each of the vectors so that a population of vectors are created where each vector may contain one of the genes from said first family and one of the genes from said  
10 second family;

c) identifying within said population of vectors a vector capable of producing a characteristic resulting from the inclusion of one gene from said first gene family and one gene from said second gene family, and using said  
15 characteristic to distinguish the vector from other vectors within the population containing other members of the gene families.

14. A method of producing a nucleic acid vector encoding two or more genes belonging to families of genes,  
20 said genes being capable of producing a characteristic that can be used to identify the vector encoding said genes from other vectors encoding other members of the families of genes, which method comprises:

a) isolating a first population of vectors, for  
25 which each member of said population may contain one member of a first family of genes and a nucleic acid site or region at which the population of vectors can be combined with a second population of vectors;

b) isolating a second population of vectors, for  
30 which each member of said population may contain one member of a second family of genes and a nucleic acid site or region at which the second population of vectors can be recombined with said first population of vectors so that one member of the first family of genes and one member of  
35 the second family of genes may be combined and expressed

in each member of a diverse population of recombined vectors;

c) recombining populations of said first and second vectors and at said nucleic acid site or region thereby  
5 creating a diverse population of recombinant vectors each of which may express one member of the first family of genes and one member of the second family of genes;

d) identifying within said population of recombinant vectors a vector capable of producing a  
10 characteristic resulting from the inclusion of one gene from each of said gene families.

15. The method of claim 14 wherein said nucleic acid site is cleaved with site-specific reagent, which method further comprises:

a) cleaving said first vector population with said reagent;

b) cleaving said second vector population with said reagent;

c) mixing both vector populations together in a  
20 ligation mix and ligating the two populations.

16. The method of claim 14 wherein said nucleic acid region is a homologous region capable of undergoing homologous recombination, which method further comprises inserting one or more members of said first and second  
25 populations into a single host capable of carrying out homologous recombination and allowing such homologous recombination to occur.

17. The method of claim 14 wherein said nucleic acid site is a target site for site-specific recombination,  
30 which method further comprises inserting one or more members of said vector populations into a single host capable of carrying out site-specific recombination at said nucleic acid site and allowing said site-specific recombination to occur.



18. The method of claim 17 wherein said target site for site-specific recombination is of the family of sites selected from flp, lox, and gamma-delta.

19. The method of any of claims 1, 11, 13 or 14  
5 wherein said vectors are plasmid or cosmid vectors.

20. The method of any of claims 1, 11, 13 or 14 wherein said vectors are phage vectors.

21. The method of any of claims 1, 13 or 14 wherein said vectors are lambda bacteriophage vectors.

10 22. The method of claim 14 wherein the identification of a particular vector within the recombinant vector population involves the interaction of sequence-specific nucleic acids with genes from said first and second families of genes.

15 23. The method of claim 14 wherein the identification of a particular vector within the recombinant vector population involves the hybridization of nucleic acid probes with genes from said first and second of families of genes.

20 24. The method of claim 14 wherein the identification of a particular vector within the recombinant vector population involves the expression of one or both of genes from said gene families as an RNA molecule.

25 25. The method of claim 14 wherein the identification of a particular vector within the recombinant vector population involves the expression of one or both of genes from said gene families as an identifiable protein molecule.

26. The method of claim 25 wherein the protein molecule(s) contains a binding site for another molecule.

27. The method of claim 26 wherein the protein molecule(s) contains an epitope recognized by an antibody.

5        28. The method of claim 27 wherein the protein molecule(s) contains an immune molecule binding site for an epitope.

29. The method of claims 14 wherein both genes express an RNA and/or polypeptide and said RNAs and/or  
10 polypeptides physically interact within a host to create said characteristic.

30. The method of claim 29 wherein both genes express polypeptides that physically interact to form a neo-epitope recognized by an immune molecule.

15        31. The method of claim 29 wherein both genes express polypeptides that physically interact to form a binding site for another molecule.

32. The method of claim 31 wherein the polypeptides are derived from antibody genes such that the interaction  
20 of both polypeptides forms an antigen binding site.

33. The method of any of claims 1, 11, 13 or 14 wherein the vectors contain a single promoter that expresses the genes from said gene families.

34. The method of any of claims 1, 11, 13 or 14  
25 wherein said genes from said gene families are each expressed from their own promoter.

35. The method of claim 11 wherein the host is a mammalian cell.

36. The method of claim 11 wherein the host is a eukaryotic cell.

37. The method of claim 11 wherein the host is a prokaryotic cell.

5        38. The method of any of claims 1, 11, 13 or 14 wherein there are more than two gene families and the vectors produced contain a random assortment of one member of each gene family needed to create said characteristic.

39. HCFLP.

10       40. LCFLP.

41. A method of producing a biological agent having a desired phenotype wherein said phenotype results from expression of a particular combined nucleotide sequence and wherein said phenotype can be used to identify the  
15 biological agent having the particular combined nucleotide sequence which comprises:

(a) bringing together a first population of nucleotide sequences with one or more other populations of nucleotide sequences to produce combined nucleotide  
20 sequences wherein each separate combined nucleotide sequence comprises one member of each population of nucleotide sequences;

(b) expressing said combined nucleotide sequences in biological agents; and

25       (c) identifying those biological agents which express said desired phenotype.

42. A method according to claim 41 wherein said phenotype can be used to distinguish the biological agent from biological agents having other combined nucleotide  
30 sequences further comprising using said phenotype to distinguish those biological agents expressing the

particular combined nucleotide sequence from biological agents having other combined nucleotide sequences.

43. A method according to claim 41 wherein said biological agent is a cell.

5        44. A method according to claim 41 wherein said biological agent is nucleic acid vector.

45. A method according to claim 41 wherein said biological agent is a bacteriophage or virus.

10       46. A method according to claim 41 wherein said phenotype results from expression of a hybrid polypeptide which is encoded by the particular combined nucleotide sequence and is encoded at least in part by one nucleotide sequence from each population of nucleotide sequences which was brought together.

15       47. A method according to claim 41 wherein said phenotype results from expression of a plurality of polypeptides wherein a polypeptide is encoded at least in part by one nucleotide sequence from each separate population of nucleotide sequences which was brought  
20 together.

48. A method according to claim 41 wherein two populations of nucleotide sequences are combined.

25       49. A method according to claim 47 wherein said phenotype results from expression of a heterodimeric polypeptide wherein one subunit of said dimer is encoded at least in part by the nucleotide sequence from the first population of nucleotide sequences and the other subunit of said dimer is encoded at least in part by the nucleic sequence from the second population of nucleotide  
30 sequences.

50. A method according to claim 48 wherein said phenotype results from expression of a first polypeptide encoded at least in part by the nucleotide sequence from the first population of nucleotide sequences and of a  
5 second polypeptide encoded at least in part by the nucleotide sequence from the second population of nucleotide sequences.

51. A method according to claim 48 wherein said phenotype results from expression of an RNA molecule  
10 encoded at least in part by the nucleotide sequence from the first population of nucleotide sequences and a second RNA molecule encoded at least in part by the nucleotide sequence from the second population of nucleotide sequences.

15 52. A method according to claim 48 wherein said phenotype results from synthesis of an RNA molecule encoded at least in part by the nucleotide sequence from the first population of nucleic acid sequences and by the nucleic acid sequence from the second population of  
20 nucleic acid sequences.

53. A method according to claim 48 wherein the first and second populations of nucleotide sequences are combined by co-infection or co-transformation of host cells.

25 54. A method according to claim 48 wherein members from said first and second populations of nucleotide sequences are combined randomly to give combined nucleotide sequences.

30 55. A method according to claim 41 wherein the combining of said populations of nucleotide sequences gives a combined nucleotide sequence which was not previously expressed in said biological agent.

56. A method according to claim 41 wherein said desired phenotype comprises a phenotype which was not previously expressed in a population of such biological agents.

5 57. A method according to claim 41 wherein said first population of nucleotide sequences comprises non-identical nucleotide sequences.

58. A method according to claim 41 wherein each population of nucleotide sequences comprises non-identical  
10 nucleotide sequences.

59. A method of producing a nucleic acid vector encoding a preselected combined nucleotide sequence which comprises two or more preselected nucleotide sequences, each independently selected from a population of nucleotide sequences, said combined nucleotide sequence being  
15 capable of producing a characteristic that can be used to identify the vector encoding said preselected combined nucleic sequence comprises

(a) bringing together a member nucleotide sequence  
20 from each population of nucleotide sequences to give a population of combined nucleotide sequences wherein each combined nucleotide sequence comprises a nucleotide sequence from each population;

(b) inserting into vector a member of the population  
25 combined nucleotide sequences so that a population of vectors is created wherein each vector may contain a combined nucleic acid sequence;

(c) identifying within said population of vectors,  
a vector capable of detectably producing a desired  
30 characteristic resulting from inclusion of the preselected combined nucleic acid sequence.

60. A method according to claim 59 wherein said characteristic can be used to distinguish the vector

encoding the preselected combined nucleotide sequence from other vectors encoding other combinations of nucleotide sequences further comprising using said characteristic to distinguish the vector from other different vectors within  
5 the population having unselected combined nucleotide sequences.

61. A method according to claim 60 wherein said nucleotide sequences are combined randomly.

62. A method according to claim 61 wherein said  
10 combined nucleotide sequences are produced using fusion polynucleotide amplification.

63. A method according to claim 59 wherein said combined nucleotide sequences are produced using fusion polynucleotide amplification.

15 64. A method according to claim 1 wherein a dicistronic or multicistronic DNA sequence which comprises one member from the first family of genes and one member from one or more than families of genes which comprises a random combination of said members of said families of  
20 genes is synthesized using fusion polynucleotide amplification and inserted into vectors.

65. A method for producing a biological agent having a desired novel phenotype wherein said phenotype results from expression of a particular combined nucleotide  
25 sequence and wherein said phenotype can be used to identify the biological agent having the particular combined nucleotide sequence; which comprises:

(a) replicating at least portions of at least two parent nucleotide sequences under conditions that allow  
30 mutations to occur in either nucleotide sequence to generate a population of diverse replicas of each parent nucleotide sequence;

(b) randomly bringing together the populations of diverse replicas to produce combined nucleotide sequences wherein each combined nucleotide sequence comprises one member of each population of diverse replicas;

5 (c) expressing said combined nucleotide sequences in biological agents; and

(d) identifying those biological agents which express said desired phenotype.

66. A method according to claim 65 wherein said  
10 desired phenotype is distinguishable from phenotypes expressed by said parent nucleotide sequences.

67. A method according to claim 66 wherein said  
phenotype can be used to distinguish it from biological  
agents having other combined nucleotide sequences using  
15 said phenotype to distinguish those biological agents  
expressing the particular combined nucleotide sequence  
from biological agents having other combined nucleotide  
sequences.

68. A method according to claim 65 wherein the  
20 parent nucleotide sequences comprise a single DNA molecule  
and are replicated together; further comprising separating  
the populations of diverse replicas of each parent  
nucleotide sequence prior to bringing together step (b).

69. A method according to claim 68 which comprises  
25 replicating two parent nucleotide sequences.

70. A method according to claim 65 wherein the  
parent nucleotide sequences are separately replicated.

71. A method according to claim 70 which comprises  
replicating two parent nucleotide sequences.



72. A method according to 71 wherein a first parent nucleotide sequence is replicated in one population of cells and a second parent nucleotide sequence is replicated in a second population of cells and said cell  
5 populations are mixed and fused to generate cells which express combined nucleotide sequences.

73. A method according to claim 72 wherein said first parent nucleotide sequences codes for a selected  $V_L$  and said second parent nucleotide sequences codes for a  
10 selected  $V_H$ , said cells are E. coli; and said combined nucleotide sequences express a Fab.

74. A method for producing a biological agent having a desired phenotype wherein said phenotype results from expression of a particular combined nucleotide sequence  
15 and wherein said phenotype can be used to identify the biological agent having the particular combined nucleotide sequence which comprises:

(a) replicating parent populations of nucleic acid sequences to generate a population of diverse replicas of  
20 each parent population:

(b) randomly bringing together the populations of diverse replicas to produce combined nucleotide sequences wherein each combined nucleotide sequence comprises one member of each population of diverse replicas;

25 (c) expressing said combined nucleotide sequences in biological agents; and

(d) identifying those biological agents which express said desired phenotype.

75. A method according to claim 74 wherein said  
30 desired phenotype is distinguishable from phenotypes expressed by said parent populations of nucleotide sequences.

76. A method according to claim 75 wherein said phenotype can be used to distinguish said biological agent from biological agents having other combined nucleotide sequences, further comprising using said phenotype to  
5 distinguish those biological agents expressing the particular combined nucleotide sequence from biological agents having other combined nucleotide sequences.

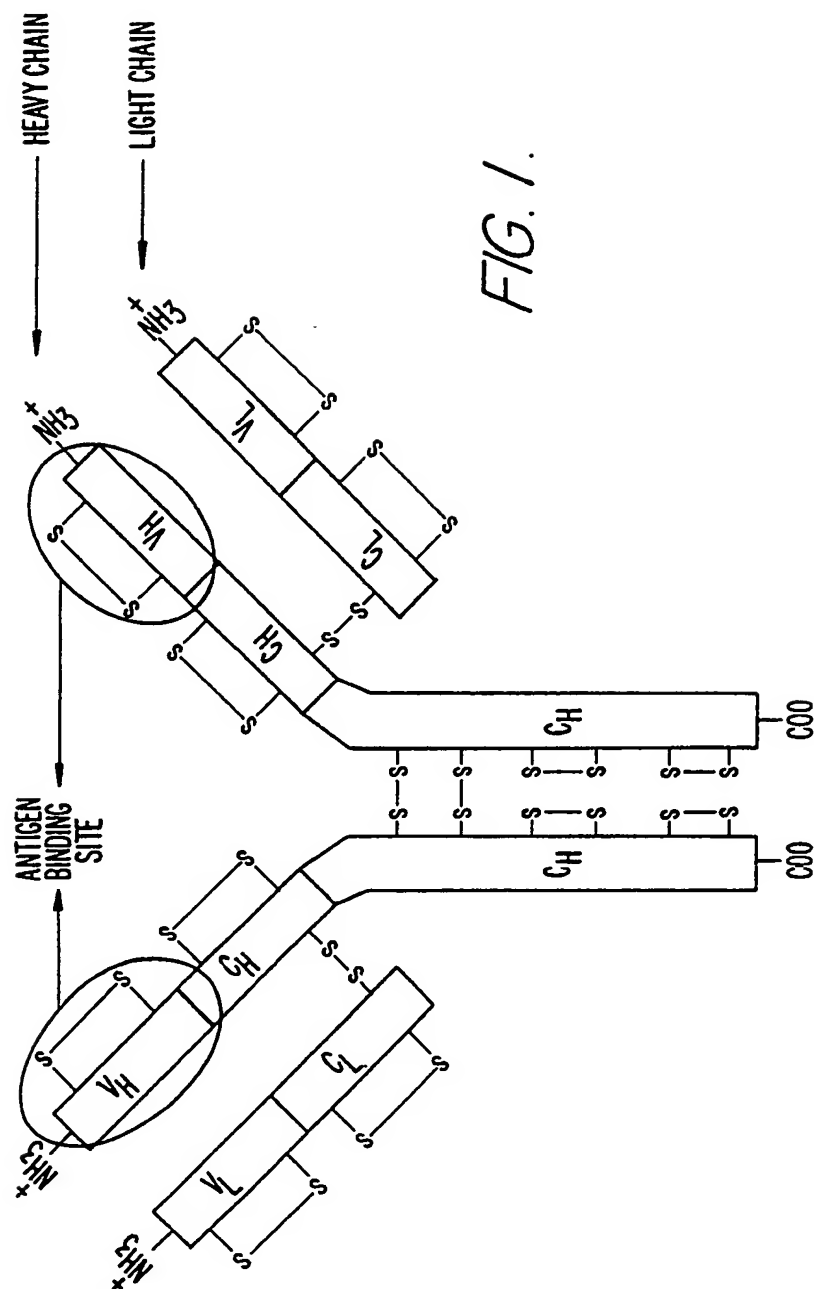


FIG. 1.

2/37

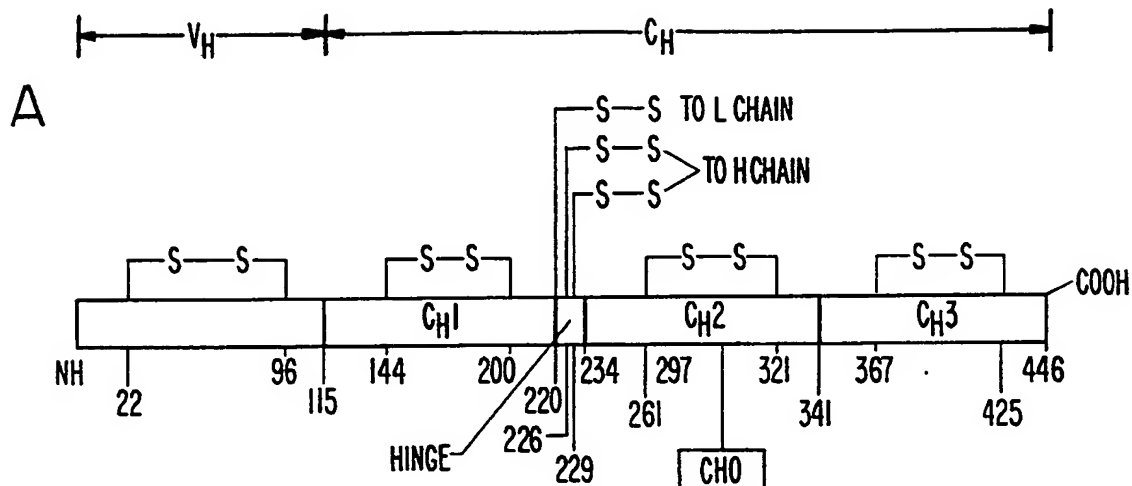


FIG. 2a.

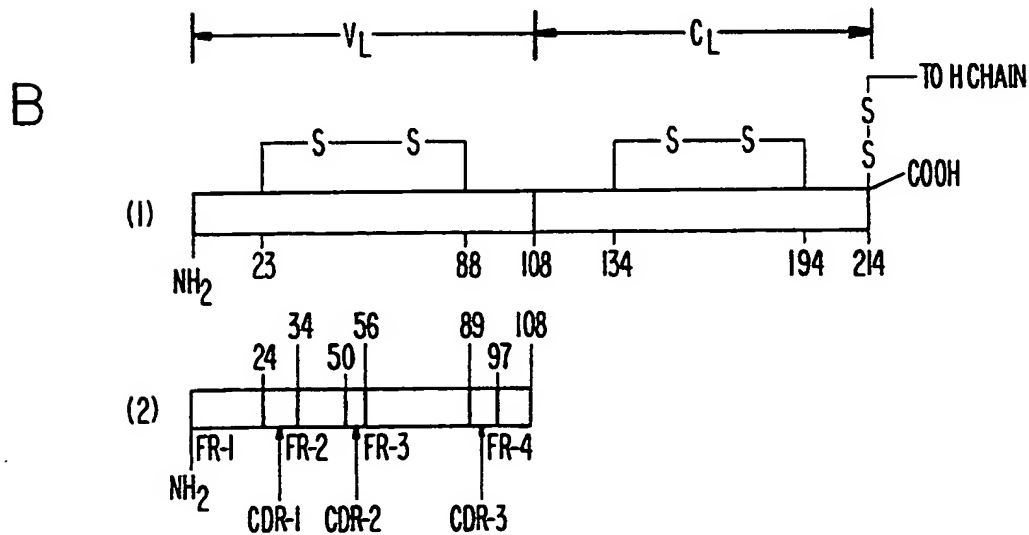


FIG. 2b.

FIG. 3-1

Class	V	
	1	10
Prototype T15	EV	KL
T15	ES	GG
	LV	Q
	PG	GS
	T15	
IgM	HPCM2	+
	HPCM3	-
	HPCM1	+
	HPCM6	-
	HPCM4	-
IgG	HPCG8	+
	HPCG13	-
	HPCG14	-
	HPCG11	+
	HPCG12	+
IgA	T15	+
	S63	+
	Y5236	+
	S107	+
	H8	+
	M603	-
	W3207	-
	M511	-
	M167	-
	V	-

FIG. 3-2

Class	V									
	20	30	40	50						
Prototype T15	LRLSCATSGTTFSDFYMEWVRQPPGKRLEWIAASRNKAN									
IgM	HPCM2									+
	HPCM3									-
	HPCM1									+
	HPCM6									-
	HPCM4									-
IgG	HPCG8									+
	HPCG13		L							-
	HPCG14			A						-
	HPCG11			I						+
	HPCG12			I						+
IgA	T15									+
	S63									+
	Y5236									+
	S107									+
	H8									+
	M603									+
	W3207									-
	M511									-
	M167									-
										-

Class		V	
IgM	Prototype T15	DYTT EYSA SVKGRFIVSRDTSQSILY LQMNALRAEDTAI	T15
	HPCM2		
	HPCM3		
	HPCM1		
	HPCM6		
IgG	HPCM4		
	HPCG8		
	HPCG13		
	HPCG14	F	
	HPCG11		
IgA	HPCG12		
	T15		
	S63		
	Y5236		
	S107		
IgG	H8		
	M603	K	
	W3207		
	M511		
	M167		
HV2		R	V
			T

FIG. 3-3

Class	V	D	J	
	100	110	120	
Prototype T15	YVCARD	YVGSS	YWYFDVWGAGTTVTVSS	T15
			Idiotype	
IgM	HPCM2			+
	HPCM3			-
	HPCM1			+
	HPCM6	DYP H		-
	HPCM4	F_RVD G		-
IgG	HPCG8	R		+
	HPCG13	A		-
	HPCG14	V_YD		-
	HPCG11			+
	HPCG12			+
IgA	T15			+
	S63			+
	Y5236			+
	S107			+
	H8	N		+
	M603	N_T		+
	W3207	N_KYD L_V		-
	M511	GD		-
	M167	T AD_N_YFG		-
		HV3		-

FIG. 3-4



7/37

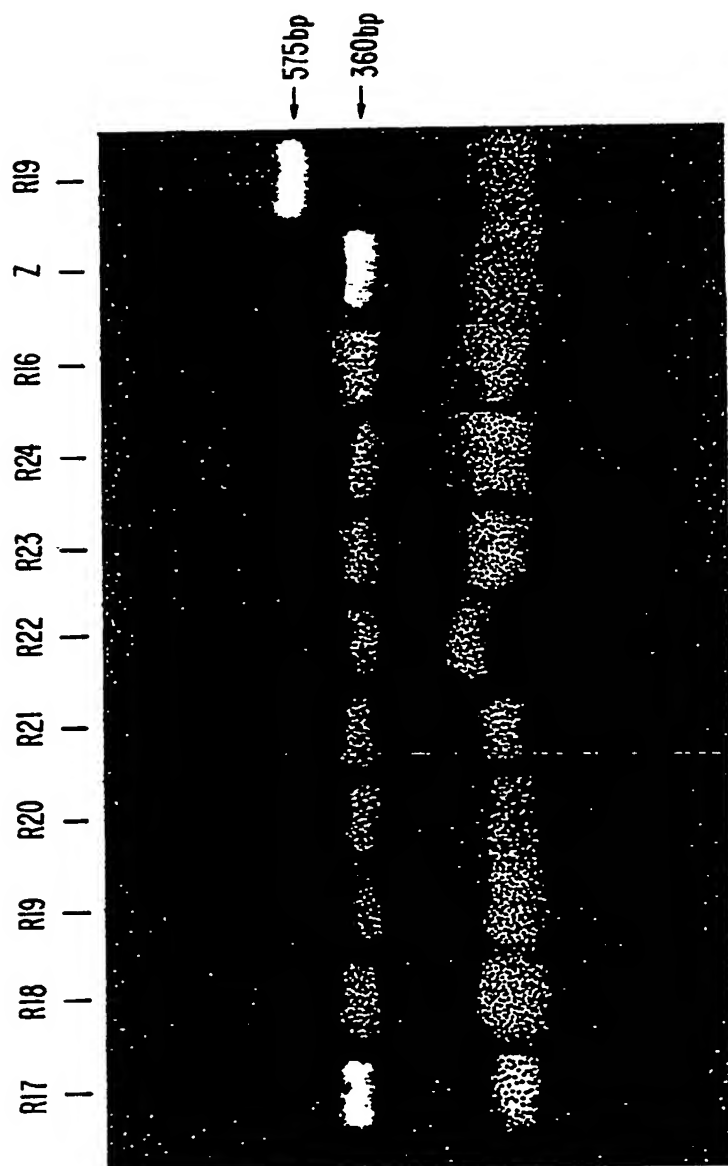


FIG. 4.

## FIG. 5-1

## Subclass I (A)

#L39

CTCGAGTCAGGACCTGGCCTCGTGAAACCTTCTCAGTCTCTGTCTCTC  
ACCTGCTCTGTCACTGGCTACTCCATCACCAGTGCTTATTACTGGAAC  
TGGATCCGGCAGTT

## Subclass II (A)

#L11

CTCGAGTCTGGGCCTnAACTGGCAAAACCTGGGGCCTCAGTGAAGATG  
TCCTGCAAGGCTTCTGGCCACACCTTGACTAGTTACTGGATACACTGG  
GTAAAAnAGAGGCC

#L03

CTCGAGTCTGGACCTnAGCTGGTAAAGCCTGGGGTTCAGTGAAGATGT  
CCTGCAAGGCTTCTGGATACACATTCACnAGCTATGTTATACACTGGG  
TGAAGCAGAAGCCT

## FIG. 5-2

#L32

CTCGAGTCTGGACCTGAACTGGTAAAGCCTGGGACTTCAGTGAAGATG  
TCCTGCAAGGCTTCTGGATACACATTCACCAGCTATGTTATGCGCTGG  
GTGAAGCAGAAGCC

## Subclass II (B)

#L37

CTCGAGTCAGGGGCTGAACTGGTGAAGCCTGGGGTTTCAGTGAAGTTG  
TCCTGCAAGGCTTCTGGCTACACCTTCACnAGCTACTATATGTACTGG  
GTGAAGCAGAGGCC

#L06

CTCGAGTCTGGGGCTAAGCTGGTAAGGCCTGGAGCTTnAGTnAAGCTG  
TCCTGnAGGGCTTCTGGCTACTCCTTCACnAGCTACTGGATGAACTGG  
GTGAAGCAGAGGCC

SUBSTITUTE SHEET

# FIG. 5-3

Subclass II (C)

#L33

CTCGAGTCTGGGGCTGAGCTGGTGAGGCCTGGAGCTTCAGTnAAGCTG  
TCCTGCAAGGCCTCTCGTACTCCTTCACCAGCTCCTGATAACTGGGTG  
AAGCAGAGGCCTGG

Subclass III (B)

#L36

CTCGAGTCAGGAGGTGGCCTGGTGAGCCTGGAGGATCCCTGAACTC  
TCCTGTGCAGCCTCAGGATTCGATTTnAGnAGATACTGGATGAATTGG  
GTCCGGCAGCTCCA

#L02

CTCGAGTCTGGAGGTGGCCTGGTGAGCCTGGAGGATCCCTGAATCTC  
CCCTGTGCAGCCTCAGGATTCGATTTnAGnAGATAATGGATGAGTTGG  
GTTCCGGCAGGCTCC

# FIG. 5-4

#L31

CTCGAGTCTGGAGGTGGCCTGGTGAGCCTGGAGGATCCCTGAAAGTC  
TCCTGTGCAGCCTCAGGATTCGATTTnAGnAGATACTGGATGAGTTGG  
GTCCGGCAGCTCCA

#L34

CTCGAGTCTGGAGGTGGCCTGGTGAGCCTGGAGGATCCCTCAAACCTC  
TCCTGTGCAGCCTCAGGATTCGATTTnAGnAGATACTGGATGAGTTGG  
GTCCGGCAGCTCCA

#L50

CTCGAGTCAGGAGGTGGCCTGGTGAGCCTGGAGGAGCCCTGAAACCTC  
TCCTGTGCAGCCTCAGGATTCGATTTnAGnAGATACTGGATGAGTTGG  
GTCCGCAGCTCCAG

## FIG. 5-5

Subclass III (C)

#L10

CTCGAGTCTGGGGGAGGCTTAGTnCAGCCTGGAGGGTCCCGGAACTC  
TCCTGTGCAGCCTCTGGATTCACTTTnAGnAGTTTGGGAATGCACTGG  
ATTCGTCAGGCTCC

#L08

CTCGAGTCTGGGGGAGGCTTAGTnnAGCCTGGAGGGTCCCGGAACTC  
TCCTGTGCAGCCTCTGGATTCACTTTnAGnAGCTTTGGGAATGCACTGG  
GTTACGTCAGGCTC

Subclass V (A)

#L38

CTCGAGTCAGGGGCTGAACTGGTGAGGCCTGGGGCGTTCAGTnAAGATG  
TCCTGCAAGGCTTCAGGCTATTCCTTCACCAGCTACTGGATGCACTGG  
GTGAAACAGAGGCC

## FIG. 5-6

Miscellaneous

#47

CTCGAGTCAGGGGCTGAACTGGCAAAACCTGGGGCCTCAGTAAAGATG  
TCCTGCAAGGCTTCTGGCTACACCTCTTCTTCTTCTGGCTGCACTGG  
ATAAAAGAAGGCCT

#L35

CTCGAGTCTGGACCTnAGCTGGTGAAGCCTGGGGTTCAGTTAAAATAT  
CCTGCAAGGCTTCTGGTACTCATTTTCTnTCTACTTTGTGAACTGGG  
TGATGCAGAGCCAT

#L48

CTCGAGTCAGGGGCTGAACTGGTGAAGCCTGGGGTTCAGTAAGTTGTC  
CTGAAGGCTTCTGGCTACACCTTCACCGGCTACTATATGTACTGGGTG  
AAGCAGAGGCCTGG

SUBSTITUTE SHEET

FIG. 6A-1

V<sub>H</sub> EXPRESSION VECTOR:

SHINE-DALGARNO MET

GGCCGCAAATTCTATTTCAAGGAGACAGTCATAATG  
CGTTTAAGATAAAGTTCCTCTGTCAGTATTAC

LEADER SEQUENCE

AAATACCTATTGCCTACGGCAGCCGCT  
TTTATGGATAACGGATGCCGTCGGCGA

LEADER SEQUENCE

GGATTGTTATTACTCGCTGCCCCAACCAG  
CCTAACAATAATGAGCGACGGGTGGTC

FIG. 6A-2

LINKER

LINKER

NCOI	V <sub>H</sub> BACKBONE	XHOI	SPEI
------	-------------------------	------	------

CCATGGCCCAGGTGAAACTGCTCGAGATTCTAGACTAGT  
GGTACCGGGTCCACTTTGACGAGCTCTAAAGATCTGATCA

STOP LINKER

TyrProTyrAspValProAspTyrAlaSer  
TACCCGTACGACGTTCCGGACTACGGTTCTTAATAGAATTCTG  
ATGGGCATGCTGCAAGGCTGATGCCAAGAATTATCTTAAGCAGCT

12/37

## FIG. 6B-1

V<sub>L</sub> EXPRESSION VECTOR:

SHINE-DALGARNO

GGCCGCAAATTCTATTTCAAGGAGACAGTCATA  
CGTTTAAGATAAAGTTCCTCTGTCAGTAT

MET

LEADER SEQUENCE

ATGAAATACCTATTGCCTACGGCAGCCGCTGGA  
TACTTTATGGATAACGGATGCCGTCGGCGACCT

LEADER SEQUENCE

TTGTTATTACTCGCTGCCCAACCAG  
AACATAATGAGCGACGGGTTGGTC

## FIG. 6B-2

LINKER

NCOI

V<sub>H</sub> BACKBONE

CCATGGCCCAGGTGAACTG  
GGTACCGGGTCCACTTTGAC

LINKER

XHOI

SPEI

STOP

CTCGAGAATTCTAGACTAGTTAATAG  
GAGCTCTTAAGATCTGATCAATTATCAGCT

SUBSTITUTE SHEET

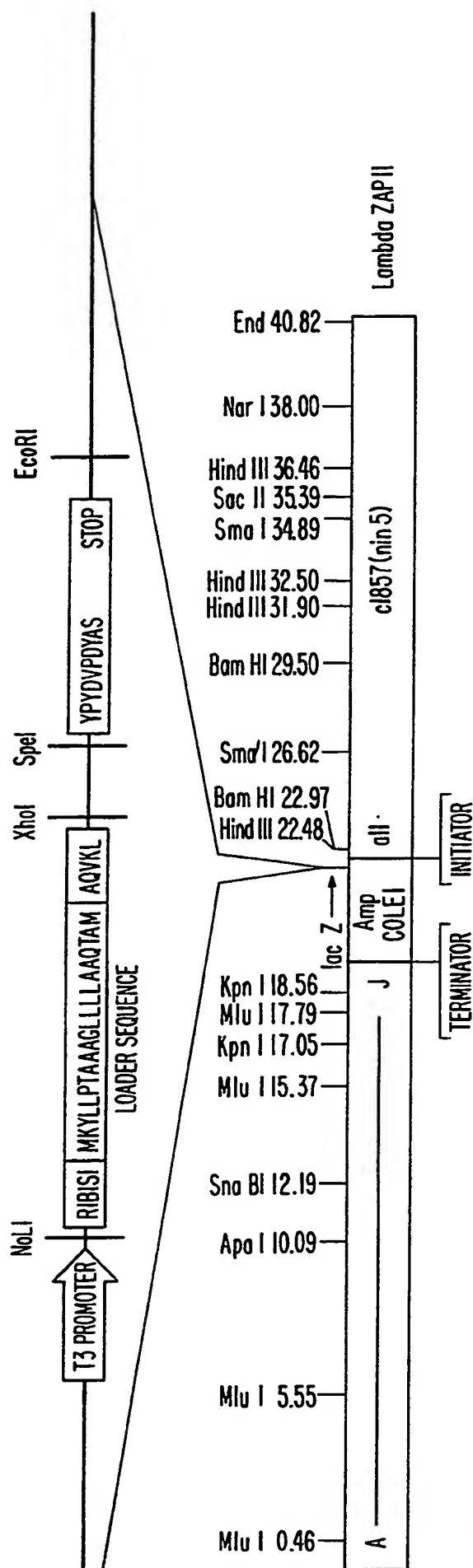


FIG. 7.

14/37

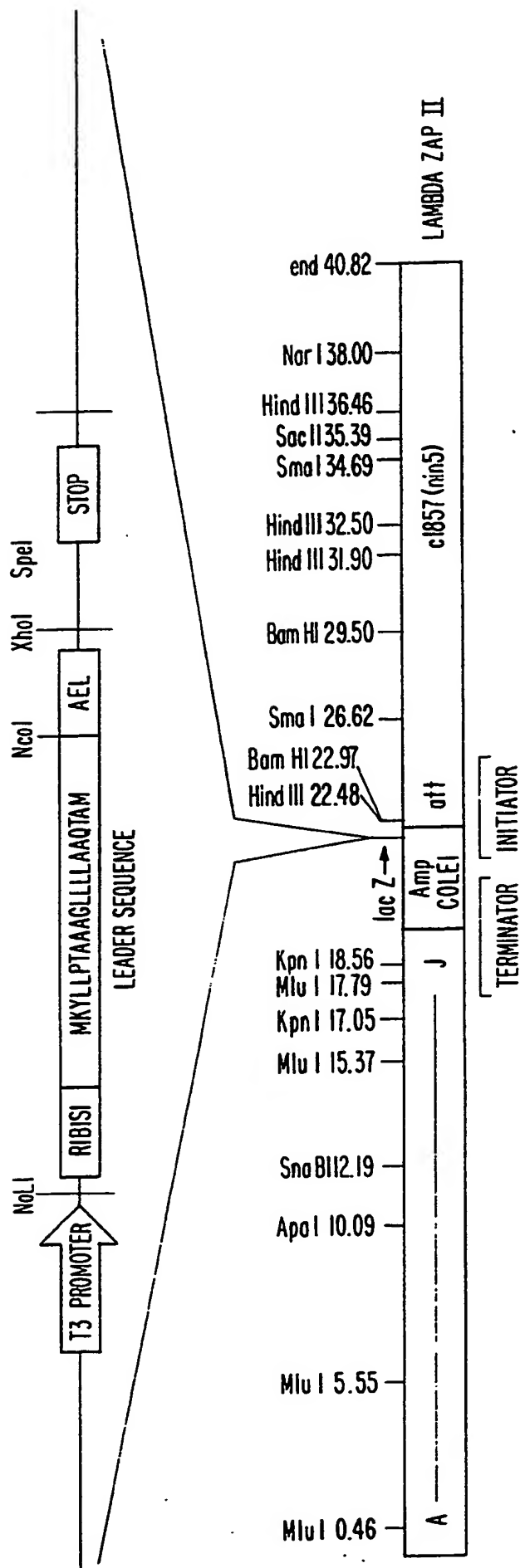


FIG. 8.



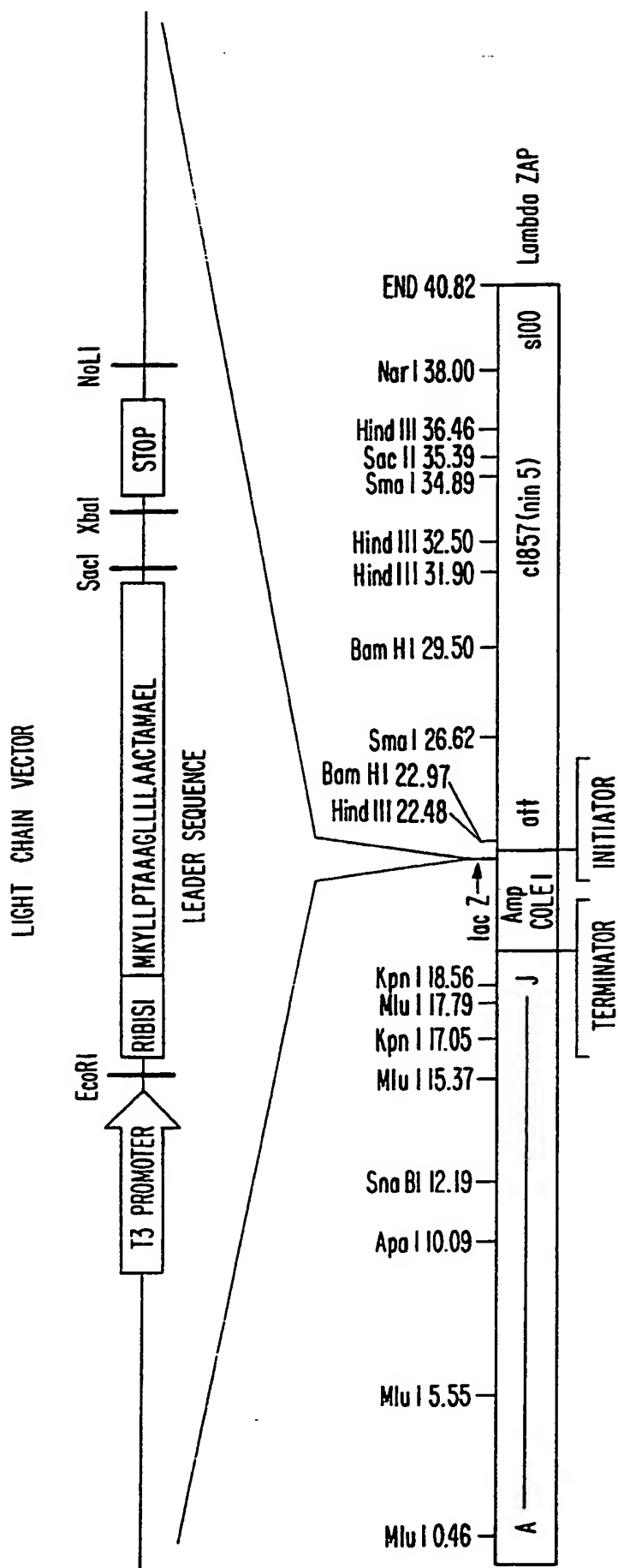


FIG. 9.

FIG. 10

ECOR I                      SHINE-DALGARNO    MET  
TGAATTCTAAACTAGTCGCCCAAGGAGACAGTCATAATGAAAT  
TCGAACTTAAGATTGTGATCAGCGGTTCTCTGTCTGTCAGTATTACTTTA

LEADER SEQUENCE

ACCTATTGCCCTACGGCAGCCGCTGGATTGTTATTACTCGCTGCCCAACCAG  
TGGATAACGGATGCCGTCGGCGACCTAACAAATAATGAGCGACGGTTGGTC

NCO I	SAC I	XBA I	Not I
<u>CCATGGCC</u> GAGCTCGTCAGT <u>TCTAGAGTTAAGCGGCCG</u>			
GGTACCGGCTCGAGCAGTCAAGATCTCAATTCCGCCGCAGCT			

FIG. 11-1

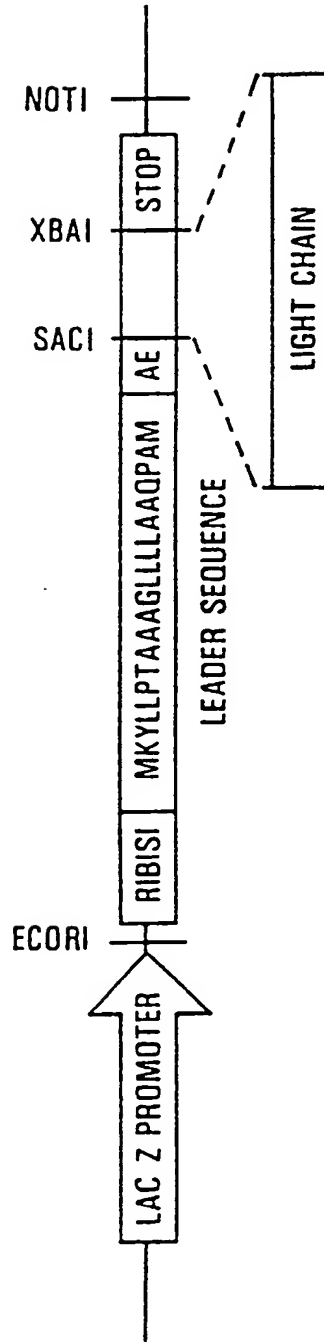
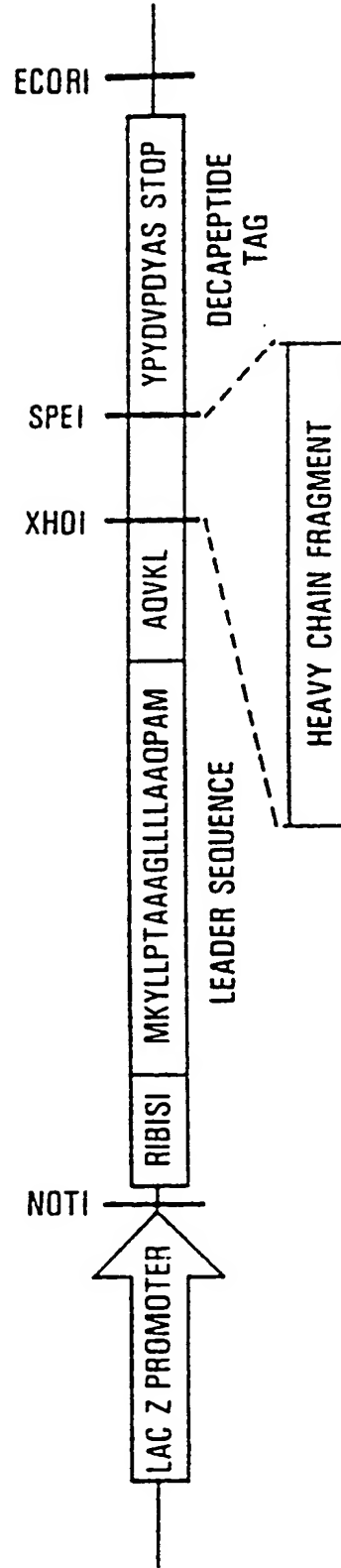


FIG. 11-2



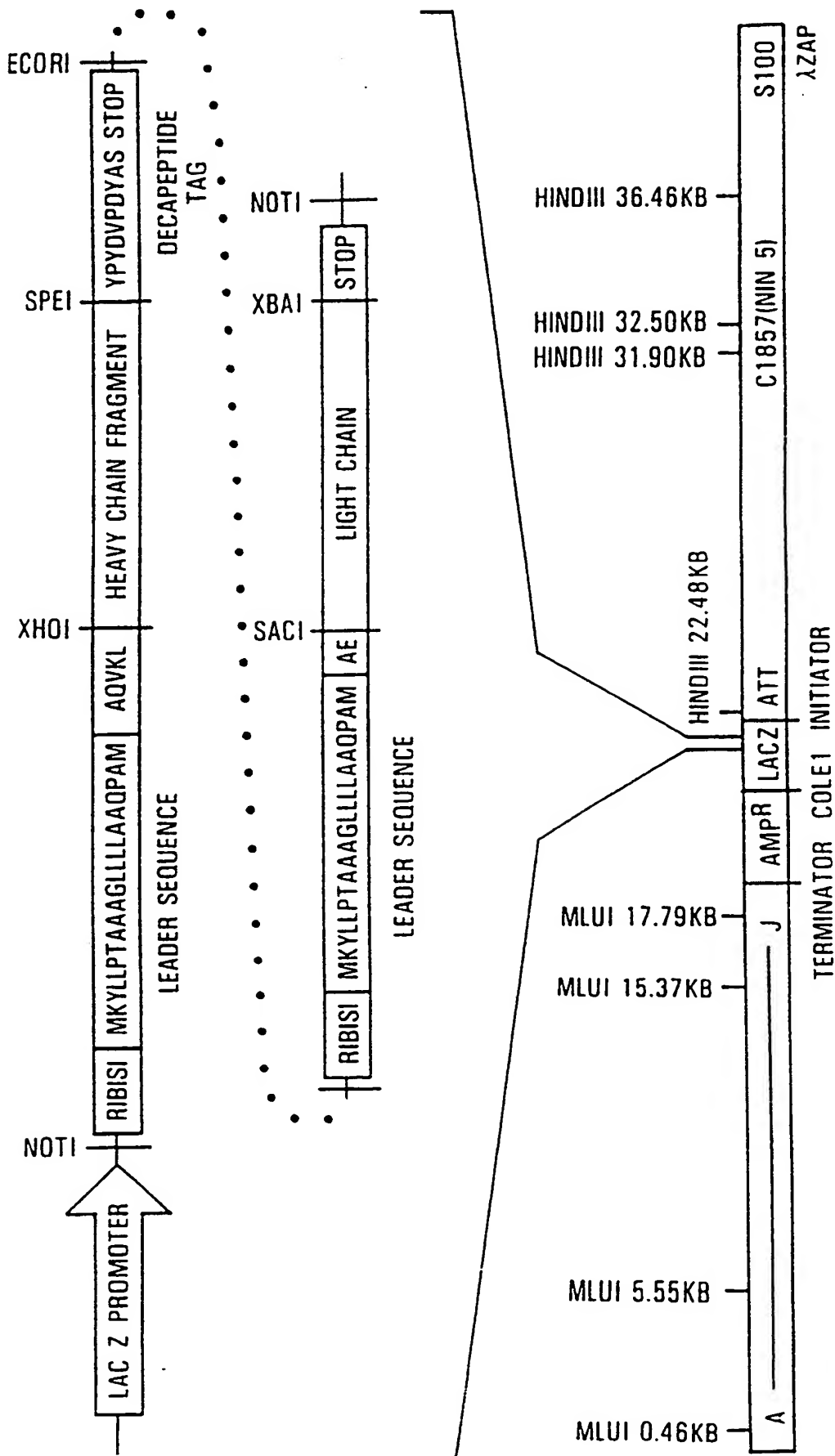


FIG. 11-3

19/37

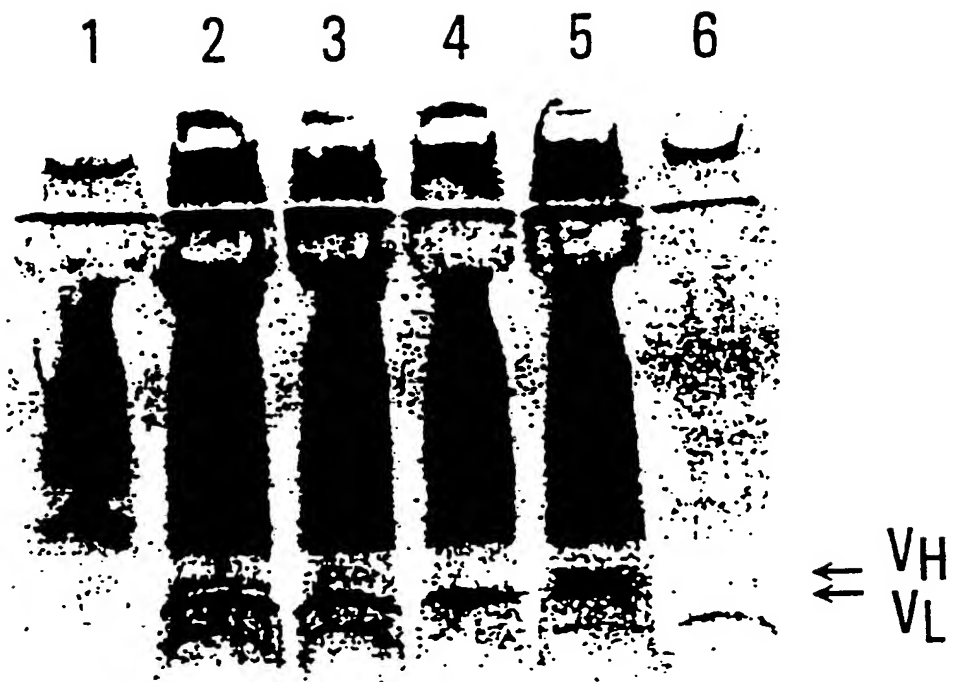


FIG. 12

20/37

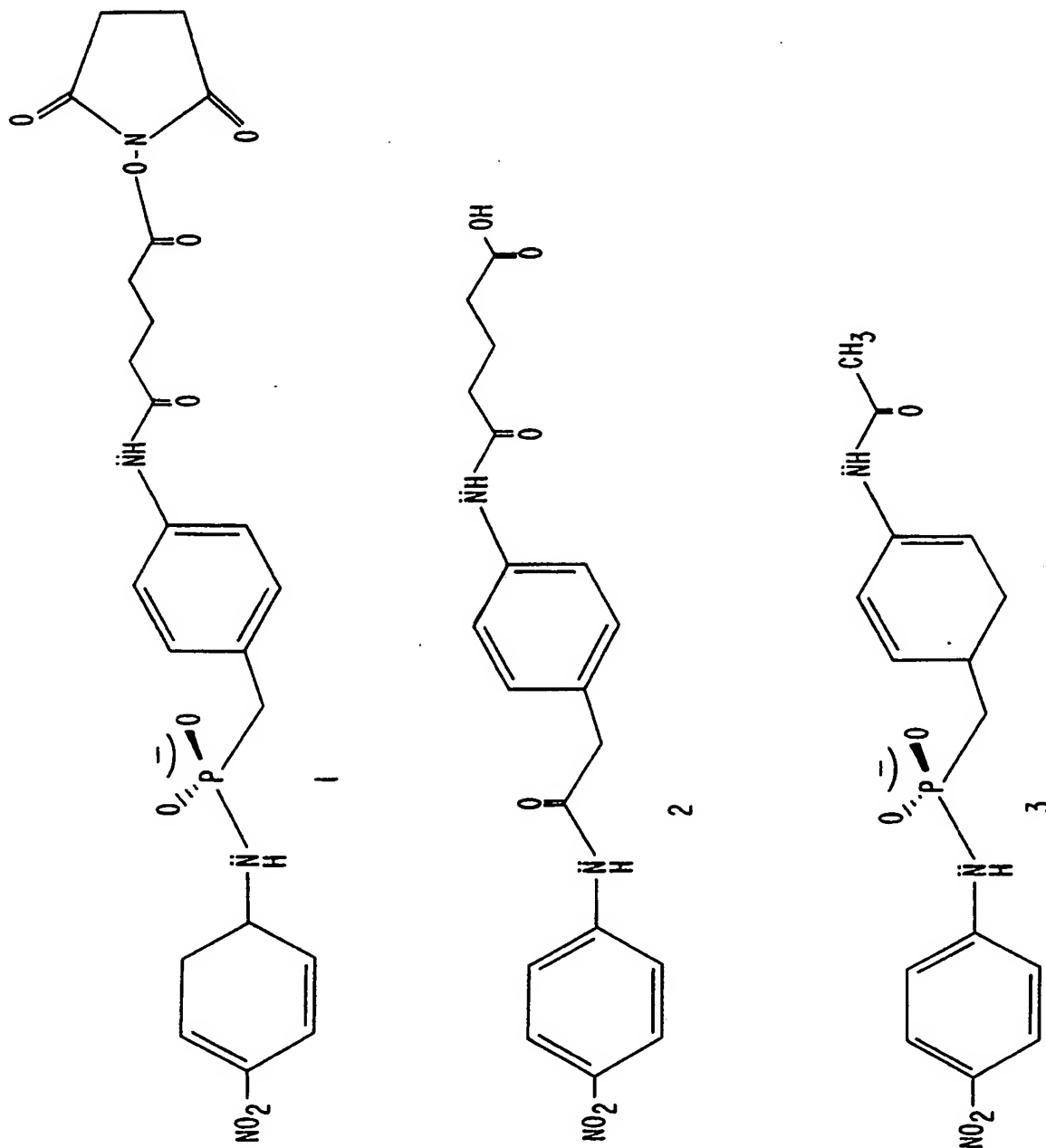
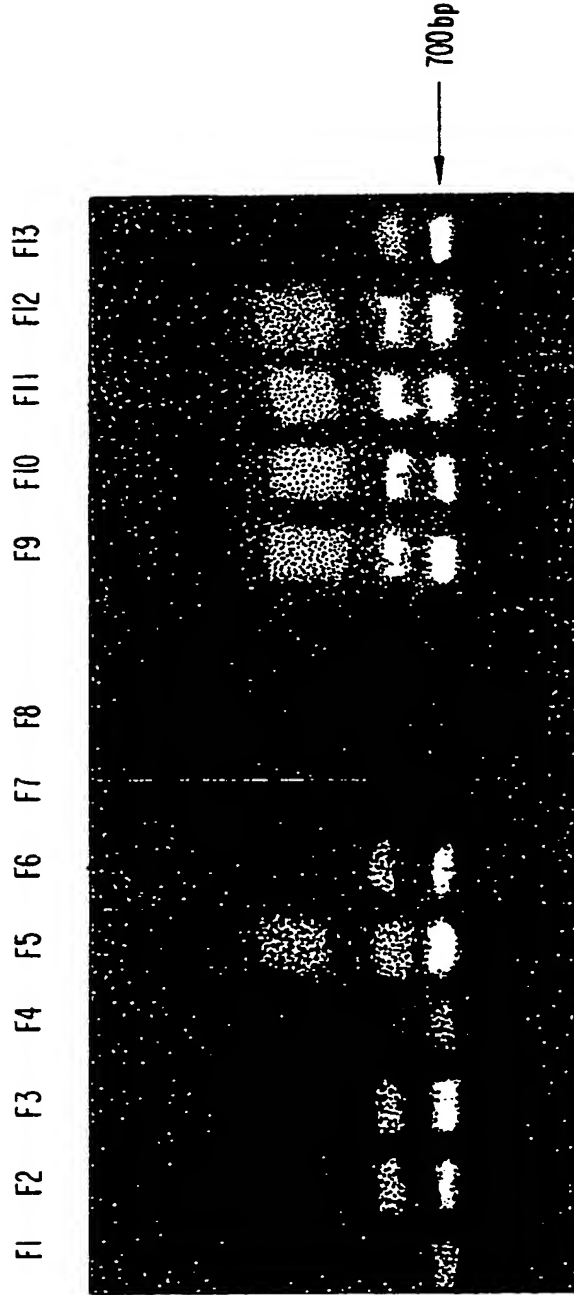


FIG. 13.

21/37

FIG. 14.



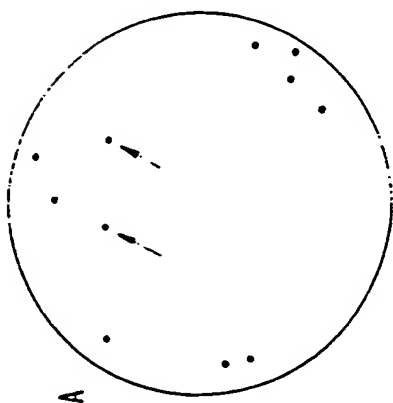


FIG. 15a.

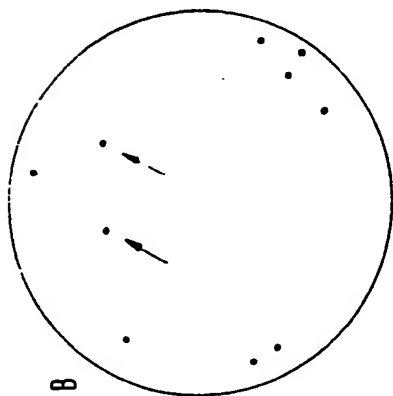


FIG. 15b.

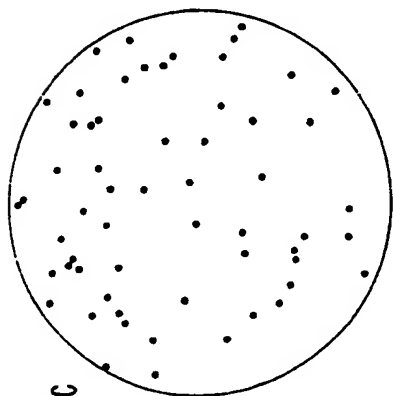


FIG. 15c.

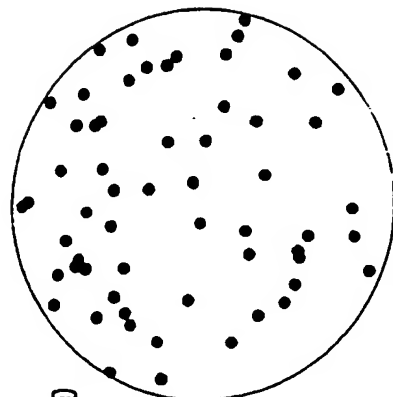


FIG. 15d.

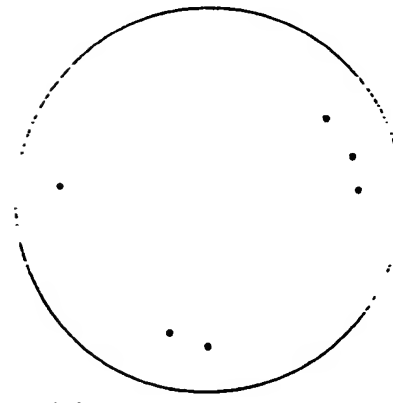


FIG. 15e.

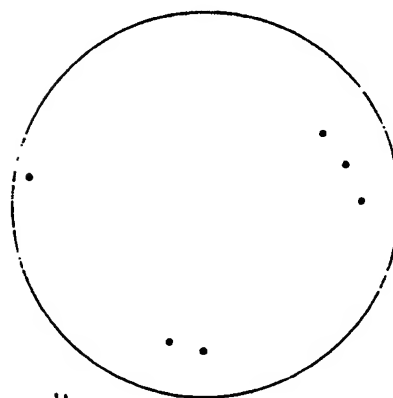


FIG. 15f.

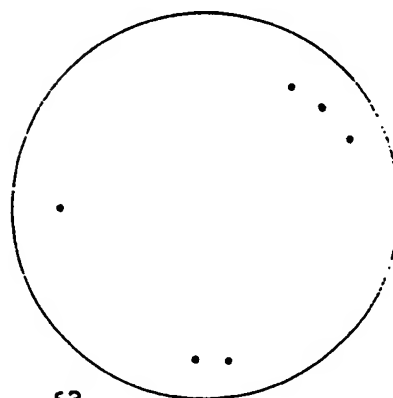


FIG. 15g.

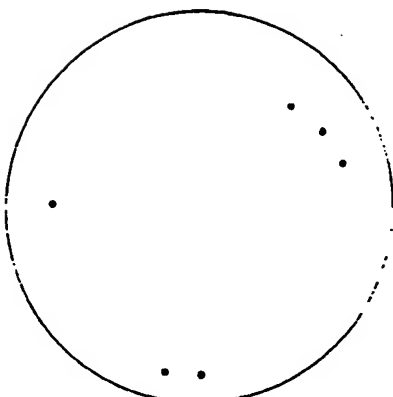
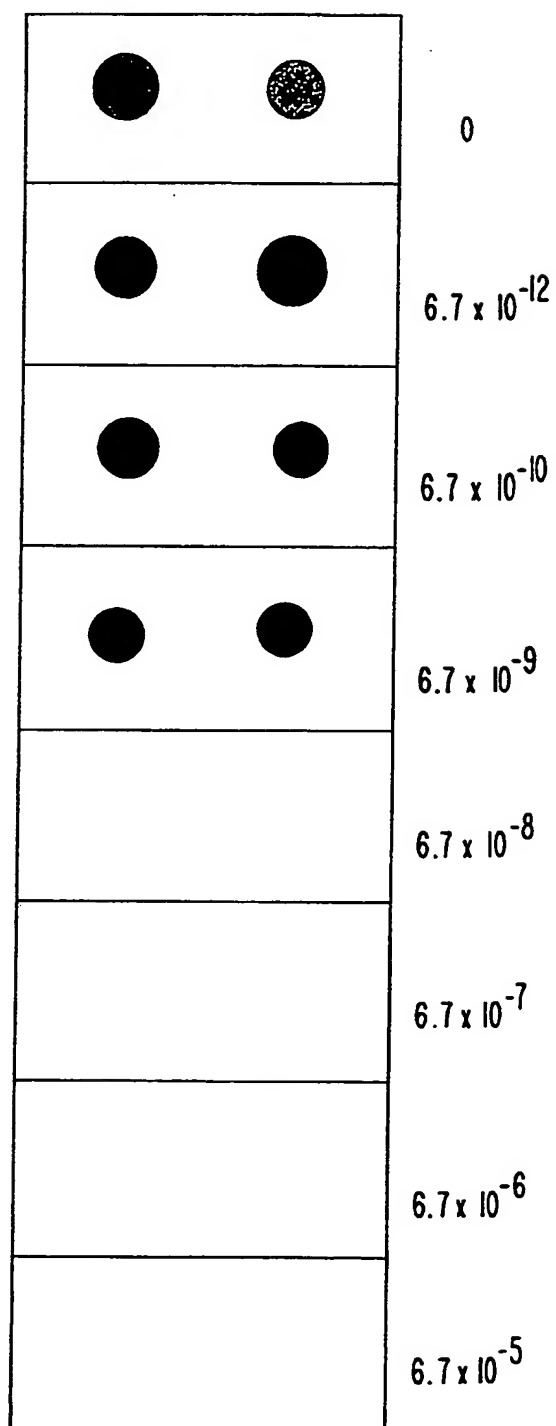


FIG. 15h.



23/37

FIG. 16.



SUBSTITUTE SHEET

24/37

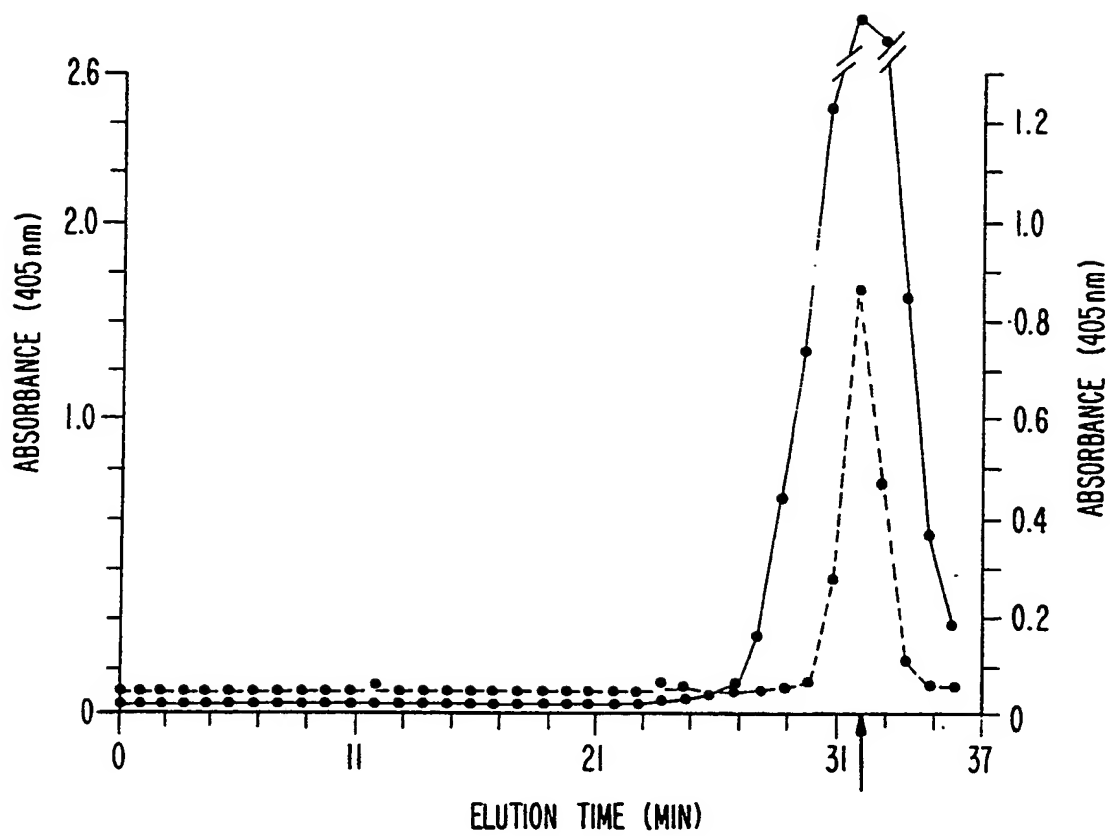


FIG. 17.

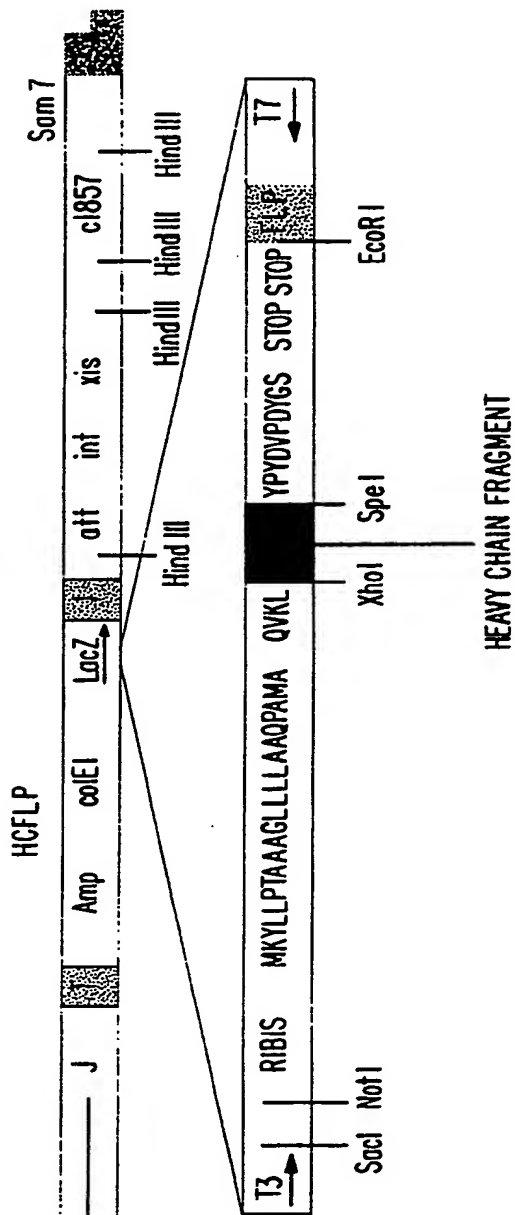


FIG. 18a.

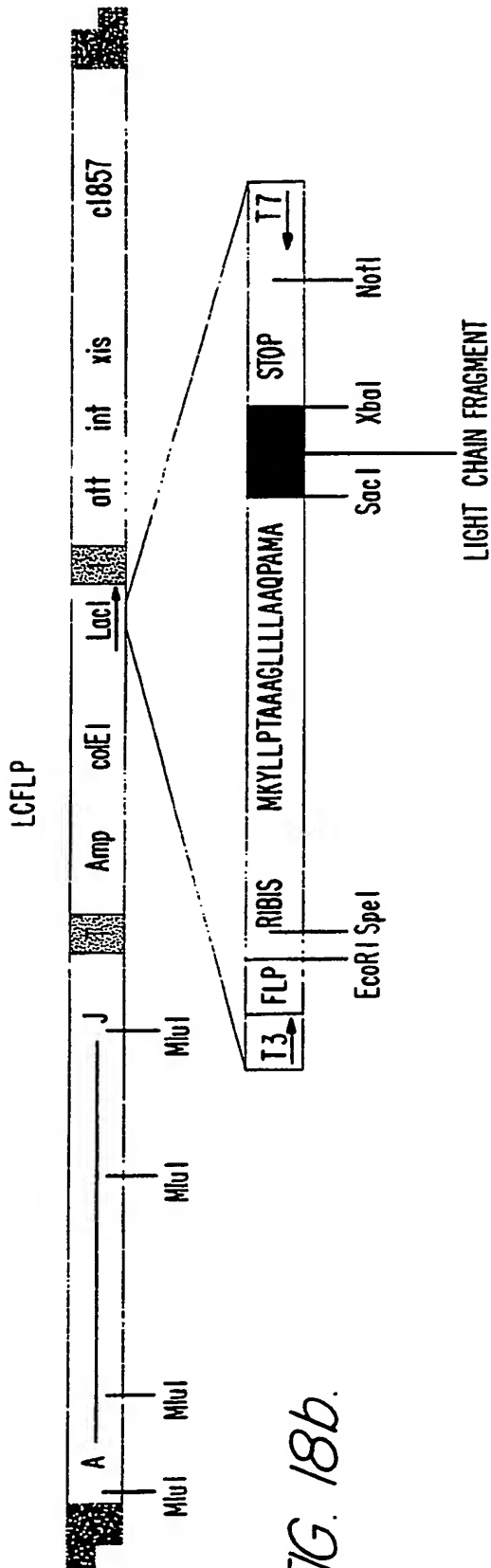


FIG. 18b.

26/37

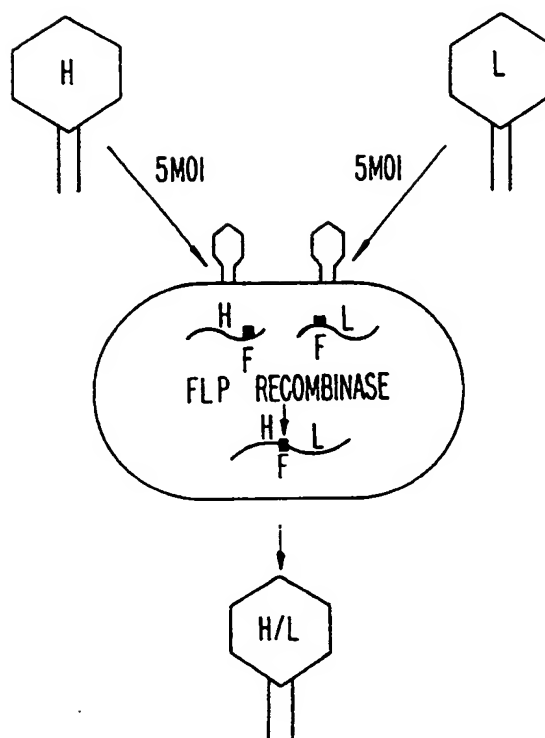
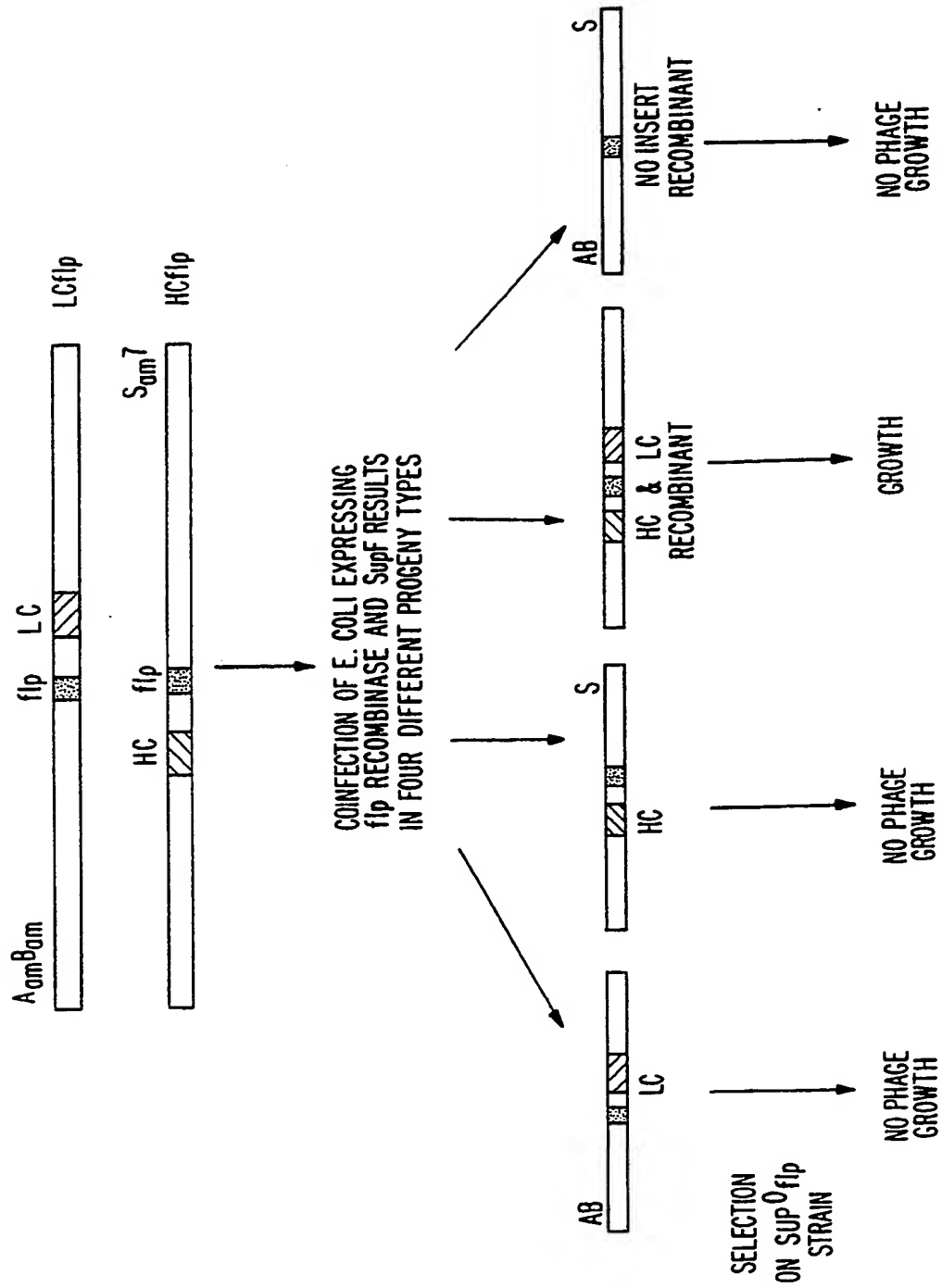
INVIVO FLP MEDIATED RECOMBINATION

FIG. 19.

27/37

FIG. 20.



28/37

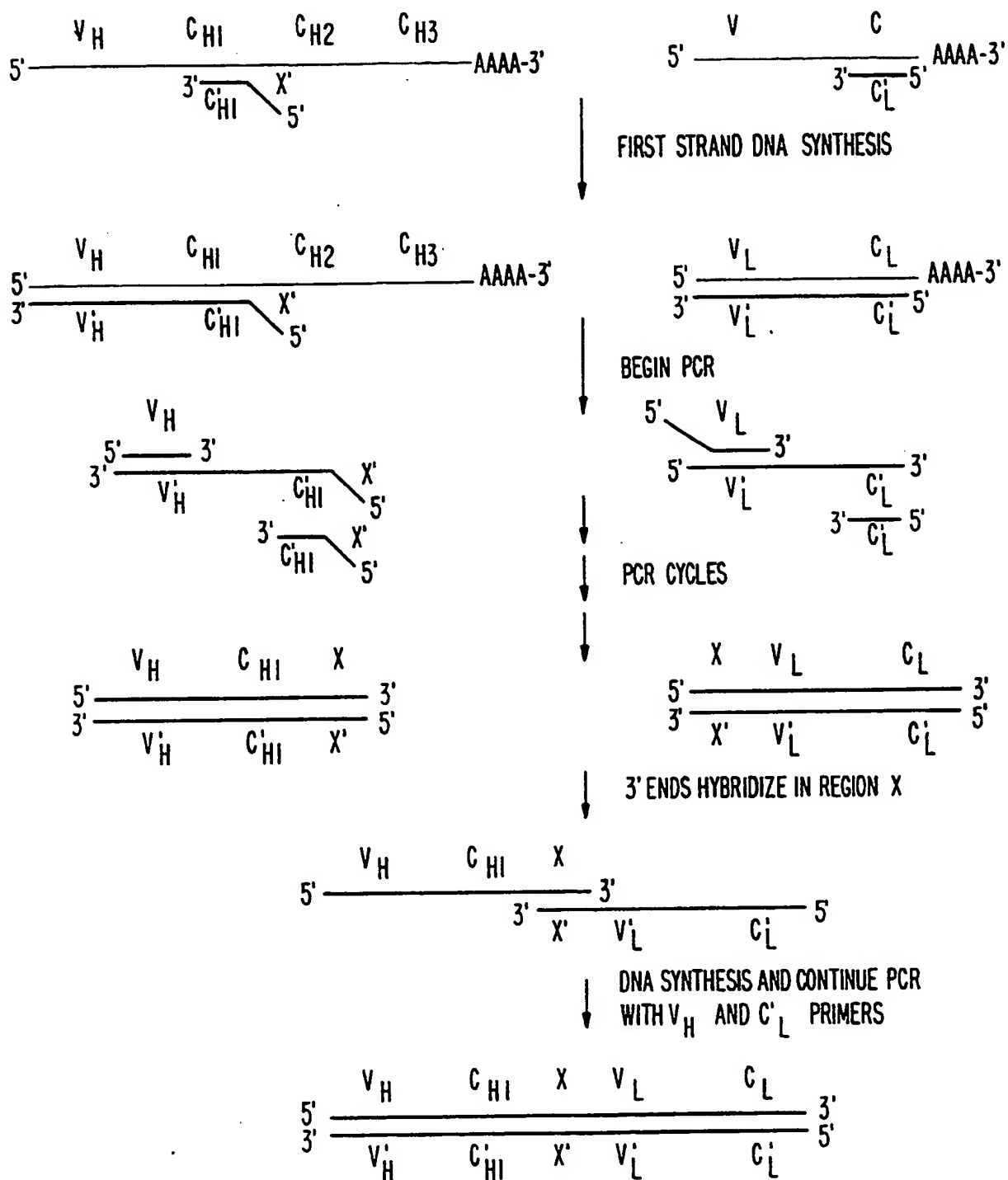


FIG. 21.

29/37

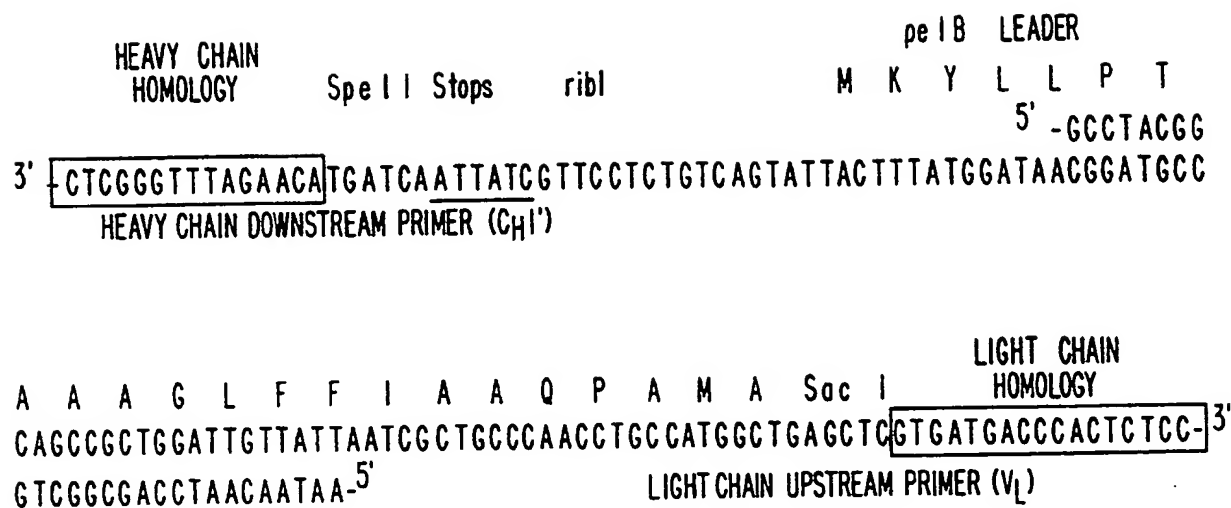


FIG. 22.

SUBSTITUTE SHEET

30/37

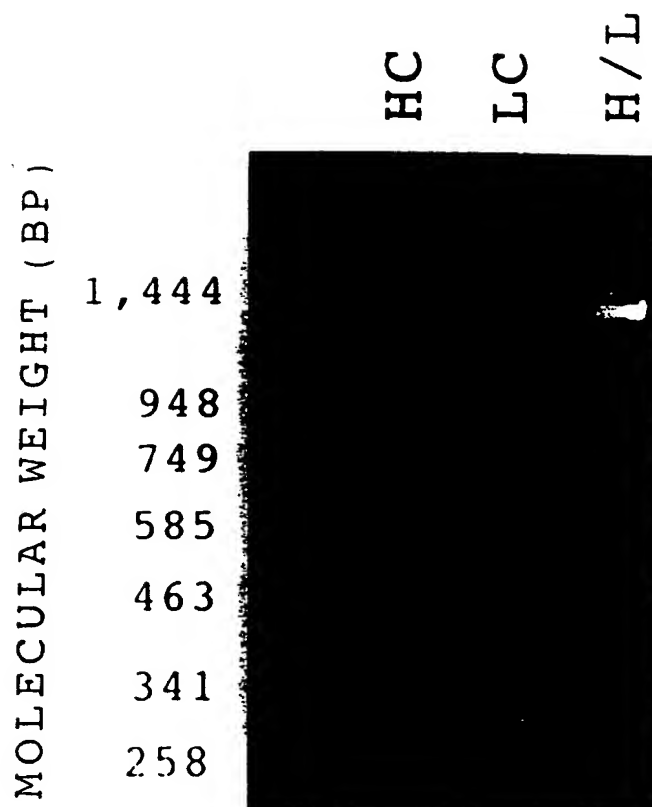


FIG. 23.

SUBSTITUTE SHEET



31/37

MODIFIED V<sub>H</sub> EXPRESSION VECTOR:

Not I      RIBOSOME BINDING SITE

5' GAGCTCGGGCCGCAAATTCTATTTCAGGAGACAGTCATA  
 3' CGCCGGCGTTTAAGATTAAAGTTCCTCTGTCAGTAT

Pel B LEADER

MetLysTyrLeuLeuProThrAlaAlaAlaGlyLeuLeuLeuLeuAla  
 ATGAAATACCTATTGCCTACGGCAGCCGCTGGATTGTTATTACTCGCT  
 TACTTTATGGATAACGGATGCCGTCGGCGACCTACAATAATGAGCGA

Nco I      Xho I      Xba I      Spe I

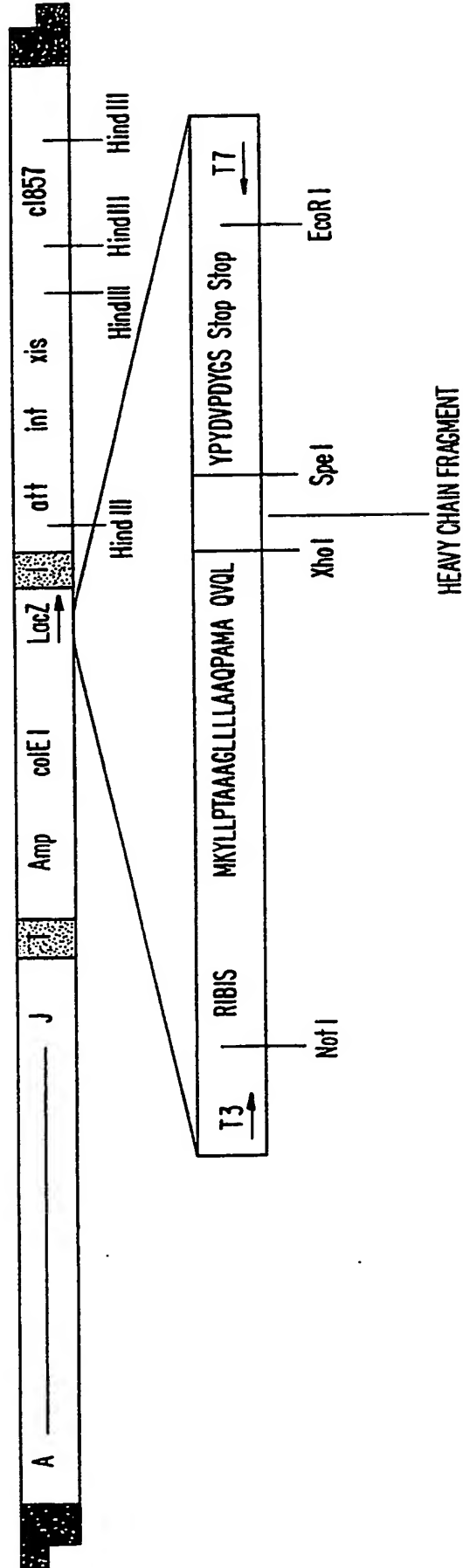
AlaGlnProAlaMetAlaGlnValGlnLeuLeuGlu      Thr  
 GCCCAACCAGCCATGCCCCAGGTGCAGCTGCTCGAGATTTCTAGACT  
 CGGGTTGGTCGGTACCGGTCACGTCGACGAGCTCTAAAGATCTGA

EcoRI

SerTyrProTyrAspValProAspTyrGlySerStop  
 AGTTACCCGTACGACGTTCCGGACTACGTTCTTAATAGAATTCG  
 TCAATGGGCATGCTGCAAGGCCTGATGCCAAGAATTATCTTAAGCAGCT

*FIG. 24a.***SUBSTITUTE SHEET**

FIG. 24b.



V<sub>H</sub> EXPRESSION VECTOR:

Not I                      RIBOSOME BINDING SITE  
5' GGGCGCAAATTCTATTTCAGGAGACAGTCATA  
CGTTTAAGATAAAGTTCCTCTGTCAGTAT

PeI B LEADER

MetLysTyrLeuLeuProThrAlaAlaAlaGlyLeuLeuLeuLeuAla  
ATGAAATACCTATTGCCACGGCAGCCGCTGGATTGTTATTACTCGCT  
TACTTTATGGATAACGGATGCCGTCGGCGACCTAACAATAATGAGCGA

NcoI                      XhoI                      XbaI SpeI

V<sub>H</sub> BACKBONE

AlaGlnProAlaMetAlaGlnValLysLeuLeuGlu                      Thr  
GCCCAACCAGCCATGCCCGAGGTGAACTGCTCGAGTTCTAGACT  
CGGTTGGTCGGTACCGGTCACCTTTGACGAGCTCTAAAGATCTGA

EcoRI

SerTyrProTyrAspValProAspTyrGlySerStop  
AGTTACCCGTACGACGTTCCGGACTACGTTCTTAATAGAATTGG  
TCAATGGGCATGCTGCAAGGCCTGATGCCAAGAATTATCTTAAGCAGCT

FIG. 25a.

34/37

V<sub>L</sub> EXPRESSION VECTOR:

EcoRI

RIBOSOME BINDING SITE

5' TCAATTCTAACTAGTCGCCAAGGAGACAGTCATA  
 3' TCGAACTTAAGATTGATCAGCGGTTCTCTGTCAGTAT

Pel B LEADER

MetLysTyrLeuLeuProThrAlaAlaAlaGlyLeuLeuLeuLeu  
 ATGAAATACCTATTGCCCTACGGCAGCCGCTGGATTGTTATTACTC  
 TACTTTATGGATAACGGATGCCGTCGGCGACCTAACAATAATGAG

*FIG. 25b.*

NcoI

SacI

AlaAlaGlnProAlaMetAlaGluLeu  
 GCTGCCCCAACCAGCCATGCGCGAGCTC  
 CGACGGGTTGGTCGGTACCGGCTCGAG

XbaI

Stop Stop

GTCAGTTCTAGAGTTAAGCGGCCG  
 CAGTCAAGATCTCAATTCGCCGGCAGCT

**SUBSTITUTE SHEET**

FIG. 26.

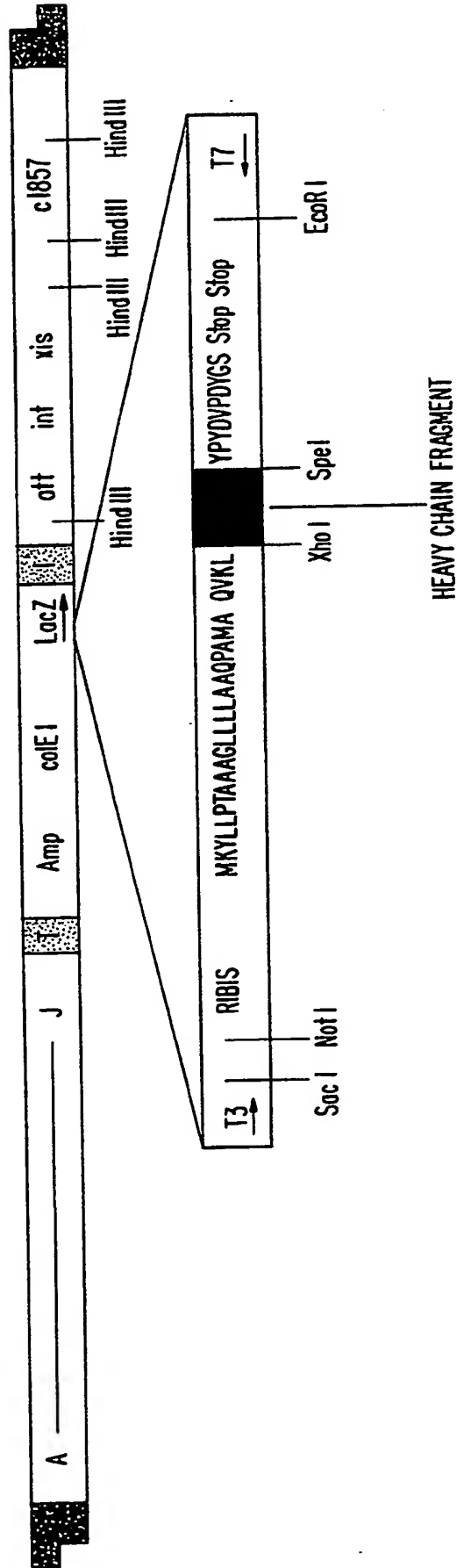
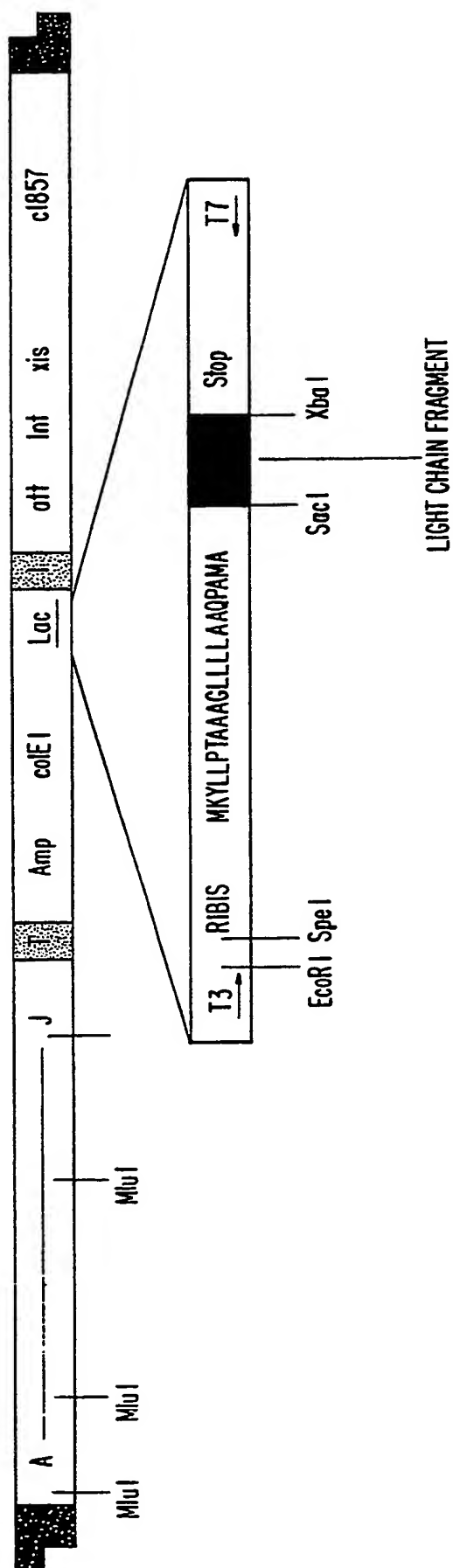


FIG. 27.



37/37

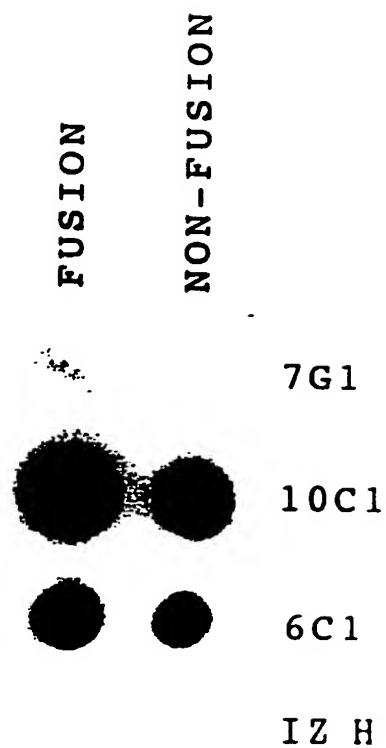



FIG. 28.

# INTERNATIONAL SEARCH REPORT

International Application No. PCT/US91/02910

<b>I. CLASSIFICATION OF SUBJECT MATTER</b> (If several classification symbols apply, indicate all) <sup>6</sup>		
According to International Patent Classification (IPC) or to both National Classification and IPC US: 435/172.1, 172.2, 172.3, 69.6, 69.7, 91; 536/27; 935/22, 23 IPC(5): C12N 15/00; C12P 19/34, 21/06; C07H 21/00		
<b>II. FIELDS SEARCHED</b>		
Minimum Documentation Searched <sup>7</sup>		
Classification System	Classification Symbols	
US	435/172.3, 91 69.6 69.7, 172.1, 172.2; 935/22, 23; 536/57	
Documentation Searched other than Minimum Documentation to the extent that such Documents are Included in the Fields Searched <sup>8</sup>		
APS		
<b>III. DOCUMENTS CONSIDERED TO BE RELEVANT <sup>9</sup></b>		
Category <sup>*</sup>	Citation of Document, <sup>11</sup> with indication, where appropriate, of the relevant passages <sup>12</sup>	Relevant to Claim No. <sup>13</sup>
Y	Proceedings of the National Academy of Sciences. Volume 82. issued November 1985. J.F. Senecoff <u>et al</u> "the FLP recombinase of the yeast 2-um plasmid: characterization of its recombination site". pages 7270-7274. See entire article.	1-10, 13 14-40 59-64
Y	US.A. 4,683,195 (Mullis et al) 28 July 1987. See entire document.	62-64
Y	US.A. 4,642,334 (Moore et al) 10 February 1987. See entire document.	1-14, 19-58.
Y	US.A. 4,816,397 (Boss et al) 28 March 1989 See entire document.	1-14, 19-58. 65-76
<div style="display: flex; justify-content: space-between;"> <div style="width: 48%;"> <p><sup>*</sup> Special categories of cited documents: <sup>10</sup></p> <p>"A" document defining the general state of the art which is not considered to be of particular relevance</p> <p>"E" earlier document but published on or after the international filing date</p> <p>"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)</p> <p>"O" document referring to an oral disclosure, use, exhibition or other means</p> <p>"P" document published prior to the international filing date but later than the priority date claimed</p> </div> <div style="width: 48%;"> <p>"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention</p> <p>"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step</p> <p>"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.</p> <p>"Z" document member of the same patent family</p> </div> </div>		
<b>IV. CERTIFICATION</b>		
Date of the Actual Completion of the International Search	Date of Mailing of this International Search Report	
24 July 1991	16 AUG 1991	
International Searching Authority	Signature of Authorized Officer	
ISA/US	 Suzanne Ziska	



**This Page is Inserted by IFW Indexing and Scanning  
Operations and is not part of the Official Record**

## **BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☒ **BLACK BORDERS**
- ☐ **IMAGE CUT OFF AT TOP, BOTTOM OR SIDES**
- ☐ **FADED TEXT OR DRAWING**
- ☐ **BLURRED OR ILLEGIBLE TEXT OR DRAWING**
- ☐ **SKEWED/SLANTED IMAGES**
- ☐ **COLOR OR BLACK AND WHITE PHOTOGRAPHS**
- ☐ **GRAY SCALE DOCUMENTS**
- ☐ **LINES OR MARKS ON ORIGINAL DOCUMENT**
- ☐ **REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY**
- ☐ **OTHER:** \_\_\_\_\_

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.**